AUTOMATED ANALYTICS FOR SLEEP DISORDERS USING MACHINE LEARNING TECHNIQUES

THAKERNG WONGSIRICHOT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCE FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2018

KMITL-2018-SC-D-002-008

AUTOMATED ANALYTICS FOR SLEEP DISORDERS USING MACHINE LEARNING TECHNIQUES

THAKERNG WONGSIRICHOT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCE FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2018
KMITL-2018-SC-D-002-008

Thesis Title	**Automated Analytics for Sleep Disorders using Machine Learning Techniques**
Student Name	Thakerng Wongsirichot
Student ID	56605014
Degree	Doctor of Philosophy (Computer Science)
Department	Computer Science
Year	2018
Thesis Advisor	Assistant Professor Dr. Anantaporn Hanskunatai

Abstract

Adults with Non-Communicable Diseases (NCD) are continuously increasing. NCDs include cardiovascular diseases, respiratory diseases, hypertension, diabetes, etc. Main causes of the NCDs are bad eating habits, daily living behaviours, lack of exercises, and sleep disorders. The sleep disorders can be in various forms including Obstructive Sleep Apnoea (OSA), which is frequently found. The Polysomnography (PSG) or the sleep test is required for sleep disorder detections. A sleep specialist is required to analyse and interpret the collected bio signals for Sleep Stage Classification (SSC) and Sleep Disorder Classification (SDC).

This research work develops two main models, Automatic Sleep Stage Classification (ASSC) and Automatic Sleep Disorders Classification (ASDC). A development of ASSC using Multi-Layer Hybrid Machine Learning Model (MLHM) is proposed. The new ASSC may improve the precisions of the analysis and interpretation of the collected bio signals. Different epoch lengths were deeply investigated. Experiments were performed in both subject-dependent and subject-independent. The MLHM achieved promising classification results, 94.20%±0.02. In addition, our proposed ASDC, opf-kNN, is a new method for classifying various types of sleep disorders. It gained the accuracy of 95.17%±3.91 with a balance between the use of high-priced medical equipment and the patient comfortability.

Keywords : Data Mining, Hybrid Machine Learning, Sleep Disorder Detection, Sleep Stage Classification

Acknowledgements

My research and thesis would have been impossible without the aid and support of my advisor, Assistant Professor Dr. Anantaporn Hanskunatai. I sincerely thank her for her continuous support during my Ph.D study. Her guidance helped me overcomes many obstacles during my study. I sincerely thank KMITL professors for their comments and suggestions during my study. This journey really lifts me up to be a stronger person and improve my visions and knowledge towards research. I would like to thank Songklanagarin Hospital for providing datasets to conduct the research. I also thank Prince of Songkla University for supporting my Ph.D. study. Last but not the least, I would like to thank my parents and those beloved one beside me for their understandings and supports during my rough and fine ways.

Table of Contents

List of Tables

List of Figures

Abbreviations/Symbols

Abbreviations/Symbols	Description
ASSC	Automatic Sleep Stage Classifications
ASDC	Automatic Sleep Disorder Classifications
AASM	American Academy of Sleep Medicine
CSA	Central Sleep Apnoea
ECG / EKG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
ICSD	International Classification of Sleep Disorders
NCD	Non-Communicable Diseases
NREM / N	Non Rapid Eye Movement
OSA	Obstructive Sleep Apnoea
PSG	Polysomnography
REM / R	Rapid Eye Movement

Chapter 1
Introduction

1.1 Background and Research Motivation

Sleep is a fundamental resting state of humans and many kinds of animals. One third of our lifetimes are allocated for sleep. During a sleep cycle, physical activities are minimised with specific operating brain activities. The Central Nervous System (CNS) plays an important role in controlling all human's activities even in the sleeping cycle. A normal sleep cycle takes approximately 7-8 hours. Sleep quality becomes one of the topics being discussed worldwide. Poor sleep quality may lead to other problems such as lower concentration during daytime, a possibility of accident, other health problems, etc. The health problems are commonly classified into a group of Non-Communicable Diseases (NCD). By definition, NCD is an abnormal health condition that is not caused by infectious agents. NCD are heart related problems, high blood pressures, diabetes, and other chronic conditions or diseases. In 2015, the World Health Organisation (WHO) reported almost 1 of 2 adults in the United States (over 133 million people) lived with at least one type of NCDs. In 2012, 17.5 million people or 31% of all worldwide was died because of the cardiovascular diseases [1]. Recently, the WHO reported approximately 40 million people deaths worldwide in 2018. The number of people with NCDs are dramatically increased annually. Specifically, from 2012 to 2018, the highest proportion of the NCD deaths remains in the category of cardiovascular diseases, which is over 17 million people [2]. In Thailand, the latest report in 2010 showed more than 400,000 NCD deaths and 29% of those were below the age of 60 [3]. The NCDs are caused by various factors such as lacks of exercises, genetic inheritances and mutations, prolonged use of some medications, etc. NCD patients' conditions are steadily degrading if medications or treatments are not properly provided. One of the known causes of the cardiovascular diseases is sleep disorders, which are initially developed from the prolonged poor sleep quality.

Sleep disorders are abnormal sleep activities, behaviours or patterns that are occurring whilst a person is sleeping. The person usually unawares of his or her abnormal unconscious behaviour. Clinically, sleep disorders can be classified into two groups namely Dyssomnias and Parasomnias [4]. A person with the Dyssomnias shows sleep problems at the beginning and during sleep period. The signs of Dyssomnias include a difficulty to sleep, some disturbances during sleep, and excessive sleepiness during daytimes. Dyssomnias can be diagnosed by quantifying a

sleep duration, an amount of disturbance times during sleep, and an overall sleep quality. One of the frequently found Dyssomnias is the sleep apnoea. Parasomnias, a less commonly found sleep disorders, refers to abnormal behaviours during sleep such as sleepwalking, sleep-eating, severe nightmares, sleep paralysis, etc. The causes of Parasomnias include genetic factors, brain disorders, etc [5].

The most encountered sleep apnoea is the Obstructive Sleep Apnoea (OSA). The OSA is occurred when the upper airway is partially or entirely blocked by collapsing soft tissue around the throat area. A patient with the OSA condition shows a series of repetitive pauses of breathing during sleep, so called apnoea. The OSA usually is developed from a number of factors including abnormal jaw structure, overweight, etc. Another less common sleep apnoea is the Central Sleep Apnoea (CSA). A patient with the CSA also shows the repetitive pauses similar to the OSA however the CSA is caused by abnormal brain activities [5]. The severity of sleep apnoea is variable, ranging from mild to severe. The sleep apnoea can be diagnosed in a Polysomnography (PSG) or a sleep test. The PSG is a gold standard procedure that collects bio-signals of a patient through a number of sensors during sleep. It is a non-invasive test. The collected bio-signals include Electrocardiography (ECG), Electroencephalography (EEG), Electromyography (EMG), and Electrooculography (EOG). Automatic Sleep Disorders Classification (ASDC) models are proposed in many research works. The intention of ASDC is to assist a sleep specialist to diagnose sleep orders from the collected PSG. However, the collected bio-signals from the PSG can be interpreted a patient's sleep stages, which is an important part of the sleep disorder diagnosis.

A sleep cycle consists of a number of sleep stages. Apart from the waking stage (W), sleep stages can be classified into two main groups namely Non Rapid Eye Movement (NREM) and Rapid Eye Movement (REM), according to the American Academy of Sleep Medicine (AASM). The NREM has three sleep stages, NREM-1, NREM-2 and NREM-3. A sleep cycle initiates when eye lids are closed (NREM-1), follows by a light sleep episode (NREM-2), and a deep sleep episode (NREM-3). The REM phrase is entered immediately after the NREM-3. The sleep stages are repetitively occurred during a sleep cycle. The sleep stages can be differentiated by the collected bio-signals from the PSG. A sleep technician is required to visually interpret the collected bio-signals and sleep disorders at the same time with limited automatic processes. The interpretation process is very time consuming. Each of the collected bio-signals is thoroughly read and interpreted [6] [7] [8]. Currently, there is no acceptable fast and efficient automatic interpretation process. Therefore, there are various groups of interdisciplinary researchers proposing alternative Automatic Sleep Stage Classifications (ASSC). In addition, Automatic Sleep Apnoea Detection

methods are also proposed. The proposed automatic methods, either ASSC or sleep apnoea detection, usually employed various knowledge such as data mining, machine learning, statistical analysis, mathematical modelling, etc. Machine learning is one of the keys to initiate alternative ways for ASSC and ASDC. Over the years, researchers attempted to propose alternative techniques both standalone and hybrid machine learning techniques. Most of the proposed techniques were constructed based on a combination of machine learning techniques, mathematical theories, and statistical models. The results are improving over the years. However, there are still some gaps that is worth to investigate.

This research work proposes new ASSC and ASDC models for classifying sleep stages and various types of sleep disorder. For other existing proposed ASSC and ASDC, the most considering part is the classification techniques especially the classification performances. However, this research work also include the concern of the minimum necessary sensors or equipment especially in the ASDC. This promotes patient comfortability, which leads to high quality sleep tests. In addition, it also benefits hospitals or medical centres which are not equipped with sophisticated PSG. For ASSC, the main goal is to classify sleep stages from various collected bio-signals. Our newly proposed ASSC is designed based on a Multi-Layer Hybrid ML Model (MLHM) that clearly shows acceptable classification results in both subject-dependent and subject-independent experiments.

1.2 Objectives of the study

1) To develop machine learning models for ASSC and ASDC.
2) To compare the developed machine learning models with ordinary ML techniques such as Decision Trees, k-Means Clustering (kMC), *k* Nearest Neighbours (*k*NN) and Support Vector Machine (SVM).

1.3 Scopes of the study

1) To study existing ASSC and ASDC models.
2) To develop new machine learning models for ASSC and ASDC.
3) To critically evaluate the new machine learning models for ASSC and ASDC compared with existing models.

1.4 Benefits of the study

1) The new machine learning models assists sleep specialist in the sleep stage classification process and the sleep disorders classification more efficiently.

Chapter 2
Theory and Literature Reviews

2.1 Sleep Disorders

Sleep disorders are abnormalities found during sleep. The abnormalities present in various behaviours. In general, we divide sleep disorders into two main groups, Dyssomnias and Parasomnias [9]. The Dyssomnias can cause sleep difficulties, sleep disturbances, and excessive sleepiness during daytimes. The abnormalities of the Dyssomnias usually occur at the beginning and during sleep period. Usually, the Dyssomnias can be diagnosed by quantifying a sleep duration, an amount of disturbance times during sleep, and an overall sleep quality. One of the mostly found Dyssomnias is the sleep apnoea, which will be discussed later. The Parasomnia include abnormal behaviours solely during sleep. The behaviours include severe nightmares, sleepwalking, sleep-eating sleep paralysis, etc [10]. The Parasomnia is caused by many factors including genetic issues, brain disorders, etc. The Parasomnia is less common than the Dyssomnias. American Academy of Sleep Medicine (AASM) is a professional society specialised in sleep medicine in the United States. AASM plays an important role in conducting research works and introducing a number of guidelines and manuals related to sleep disorders and sleep stage classifications. Recently, AASM announced the third edition of International Classification of Sleep Disorders (ICSD). ICSD is a document that covers the classification of sleep disorder internationally [11].

Common sleep disorders can develop to more serious sleep disorders that is sleep apnoea. Sleep apnoea is a serious condition occurred during sleep. A person with sleep apnoea shows repeatedly stopped breathing. Obstructive Sleep Apnoea (OSA) is one of the most encountered sleep apnoea. The OSA presents a condition of obstruction of the upper airway with series of repetitive pauses of breathing (apnoea) during sleep. Causes of the OSA are overweight, abnormal jaw structure, etc. A less common type of sleep disorder is the Central Sleep Apnoea (CSA). It has some similarities to the OSA but the CSA is caused by some abnormalities in the central brain [11] [12]. Both OSA and CSA are usually found in severe patient cases.

In addition, there are other frequently found sleep disorders and symptoms that can be found together with the above sleep disorders. For example, Oxygen Desaturation, Hypopnea, Isolated Limb Movement, and Periodic Limb Movement. According to AASM guidelines, an oxygen desaturation event is counted when a present of 3 percent reduction of blood oxygen level during any respiratory event. In

common, the oxygen desaturation is measured by oxygen desaturation index (ODI), a number oxygen desaturation event detected per hour. Hypopnea is classified into a type of sleep apnoea. The hypopnea shows a certain degree of breathing blockage whilst the sleep apnoea is a full breathing blockage. The level of severity are measured by Apnea Hypopnea Index (AHI). The AHI is calculated by counting a number of hypopnea and apnoea found per hour. The AHI over 30 is considered to be severe. Isolated Limb Movement is a short segmental movement of the lower or upper limbs during sleep. Periodic Limb Movement is a series of rhythmic and repetitive leg movement, usually happening and last within 20-40 seconds [8] [11].

2.2 Polysomnography (PSG)

Polysomnography (PSG) or a sleep test is a gold standard test for studying sleep patterns and identifying sleep abnormalities. In the old days, the PSG was conducted in a paper-based machine. The paper-based PSG was conducted in 30 centimetres page of papers. The paper speed was 10 millimetre per second. Therefore, one epoch contained 30 seconds of signal data based on the physical limitation of papers. The recent digital PSG carried on the fashion of 30 seconds epoch length. In addition, the digital PSG allows us to display signals in various time lengths including 5, 10, 30, 60, 90, 120 and 240 seconds. It helps sleep specialists to mimic details of particular signal and also look for broader view in longer epoch lengths.

Electroencephalogram (EEG) plays an important role in brain pattern recordings in order to identify abnormal brain waves. It assists to diagnose of most brain related diseases. The EEG operates by measuring voltage differences, so called derivations, between two corresponding electrodes. In practice, the electrodes are placed in different locations in the standard EEG as shown in Figure 2.1. The AASM recommends selective electrodes in the standard EEG recording for the PSG. In PSG, Frontal (F), Central (C) and Occipital (O) electrodes are used with referencing with the opposite mastoid (A or M) electrodes. Specifically, there are three sets of derivations recommended in the ASSM scoring manual namely F4-M1, C3-M1, and O2-M1. The backup derivations include F3-M2, C3-M2, and C1-M2, respectively [6] [8].

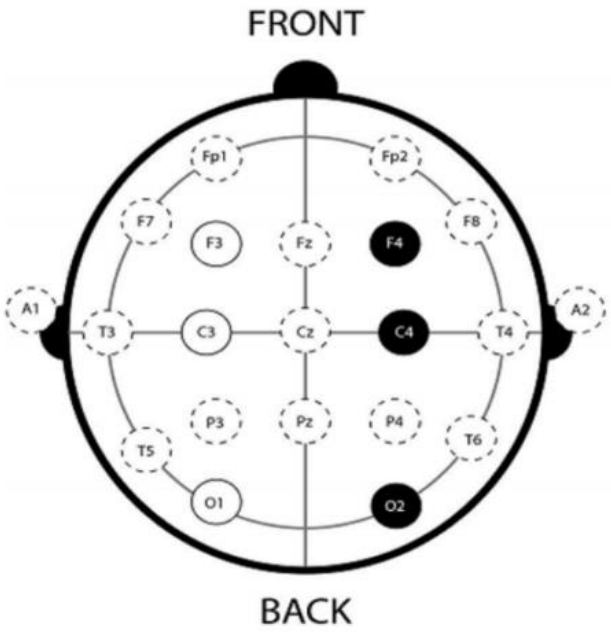


Figure 2.1 EEG electrode placements

Sleep specialists visualise EEG waveforms in order to identify sleep stages and in some cases localise sleep apnoea. Common EEG frequency are delta (< 4 Hz), theta (4-7 Hz), alpha (8-13 Hz) and beta (>13 Hz). Other special patterns include “Sharp waves”, which are narrow wave pattern within 70-200 msec, “spikes”, which have shorter interval of 20-70 msec. In terms of sleep stage classification, six EEG patterns are commonly taken into account. Firstly, “Alpha rhythm” or the alpha wave is an oscillation wave form with a frequency between 8-13 Hz. The Alpha rhythm activity is usually increased when eyes closed and decreased when eyes opened. “Sleep spindle” is presented in spindle-shaped wave form with frequency between 11-16 Hz. The sleep spindles are high in the central area. Some drugs may cause the sleep spindles faster. “K complex” is an extremely high amplitude wave form that continuously shows both high negative and positive wave amplitudes. The K complexes are peak in the frontal area. “Vertex sharp wave” is a shape wave form with a very short duration (<500 milliseconds). The vertex sharp waves are visible in the central area. They are sharply outstanding from the background activity. “Slow wave” activity is a high amplitude wave form with a frequency between 0.5-2 Hz. The slow wave activities are usually transmitted to the derivations that relates to eyes. “Saw-tooth wave” is a triangular or saw-like wave form with a frequency between 2-6 Hz. The saw-tooth waves are usually found in the central area. Table 2.1 and 2.2 show a sample of EEG wave in various patterns discussed above.

Table 2.1 A sample of 15 seconds EEG wave for Alpha Wave, Sleep Spindle and K Complex

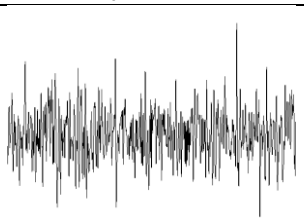
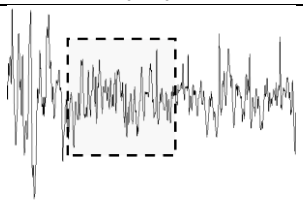
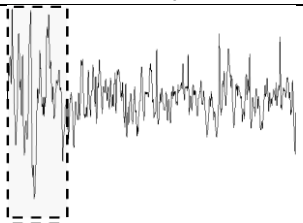
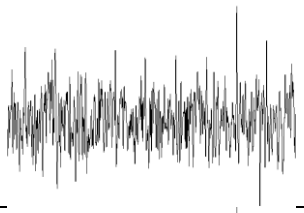
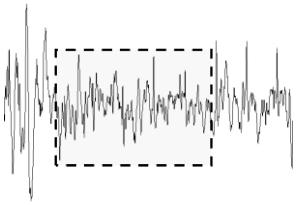
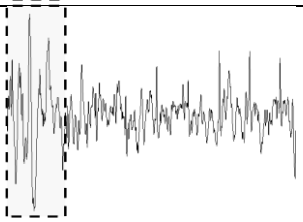
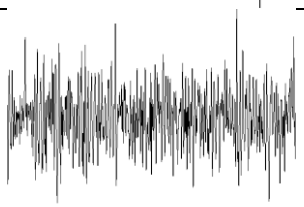
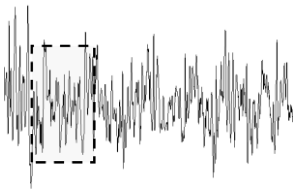
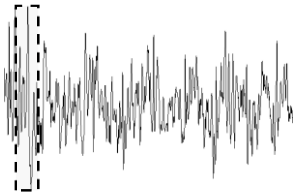
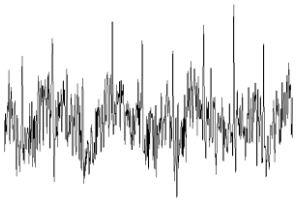
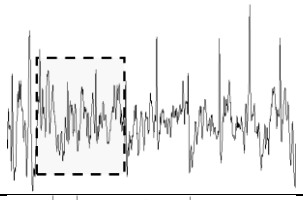
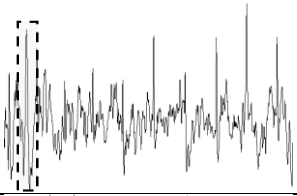
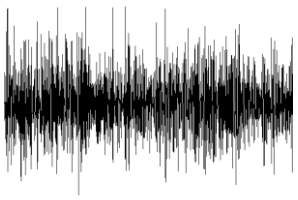
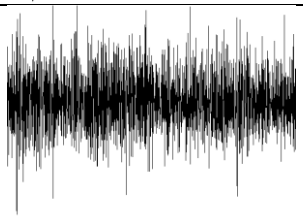
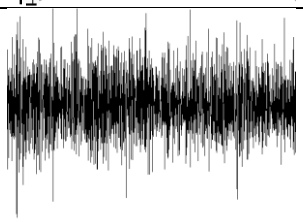
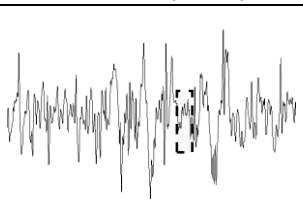
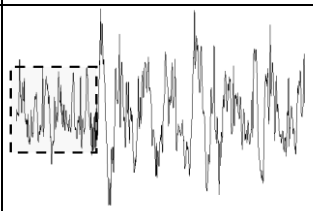
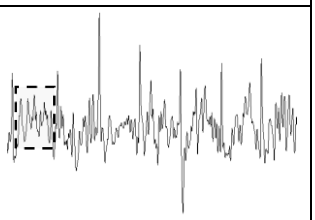
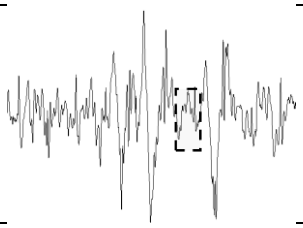
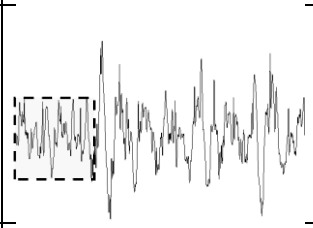
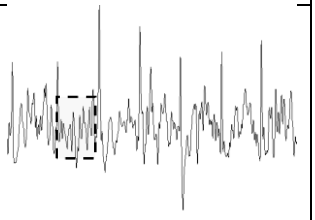
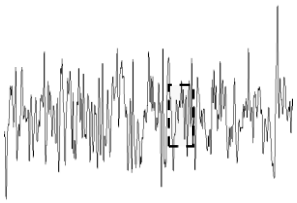
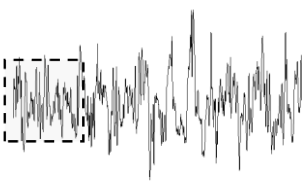
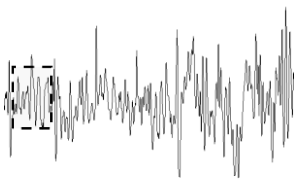
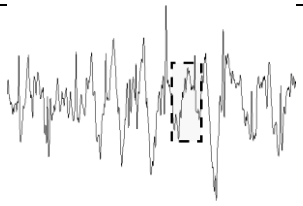
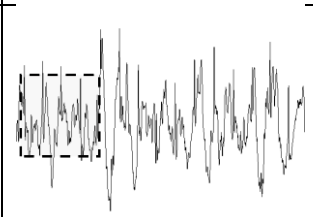
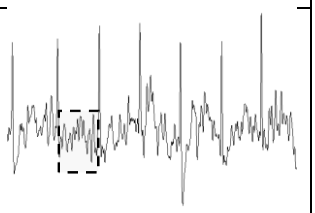
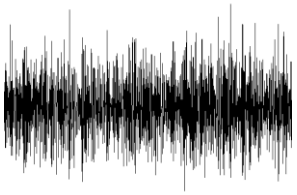
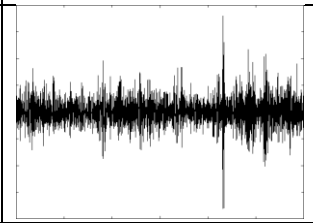
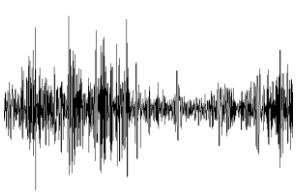
Signal	Alpha Wave	Sleep Spindle	K Complex
C4_A1			
F4_A1			
O2_A1			
ROC_A1			
CHIN			

Table 2.2 A sample of 15 seconds EEG wave for Vertex Sharp Shape, Slow Wave and Saw-tooth

Signal	Vertex Sharp Shape	Slow Wave	Saw-tooth
C4_A1			
F4_A1			
O2_A1			
ROC_A1			
CHIN			

Electrooculography (EOG) collects voltage differences between eye electrodes. The AASM recommends to place two electrodes, E1 and E2, as shown in Figure 2.2. The E1 and E2 are previously known as LOC and ROC, in the left and right eyes, respectively. The voltage differences are measured in the montage between E1-M2 and E2-M2. If eyes are moved toward an electrode, the positive value of voltage changes are shown. In the opposite side, the voltage changes are shown as negative.

Common EOG patterns include eye blinks, reading eye movements, slow eye movements (SEMs), and REMs. Eye blinks are vertical eye movements within a frequency of 0.5 – 2 Hz. Reading eye movements shown series of a slow phase followed by rapid phase in the opposite direction. Slow eye movements are

sinusoidal eye movements for more than 500 msec. Rapid eye movements consist of irregular sharp eye movements that can be last for less than 500 msec.

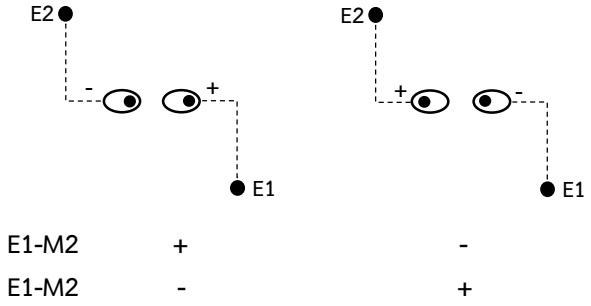


Figure 2.2 E1-M2 and E2-M2 electrode placements

Electromyography (EMG) is a procedure to evaluate specific muscles and nerve cells. EMG results can show muscle and nerve dysfunction. In the PSG, mostly the Chin-EMG, the electrodes attached to the chin area, is used to identify the deep sleep stage, so called REM stage. In the deep sleep stage, minimum Chin-EMG activities are being visualised. The Chin-EMG activities are lower than the minimum Chin-EMG activities found in other sleep stages. AASM recommends to locate three Chin-EMG electrodes in the chin and neck area. The first electrode is placed in the centre of the chin. The rest two electrodes are located in both sides of the neck area as shown in Figure 2.3. Two montage derivations are measured including Chin2-Chin1 and Chin3-Chin1. The rest of the electrodes, which are not related to Chin-EMG, are used as alternative electrodes. Table 2.3 shows a sample of 30 sec Chin-EMG activities.

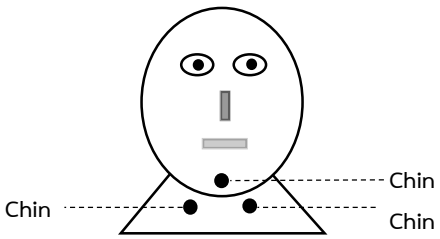
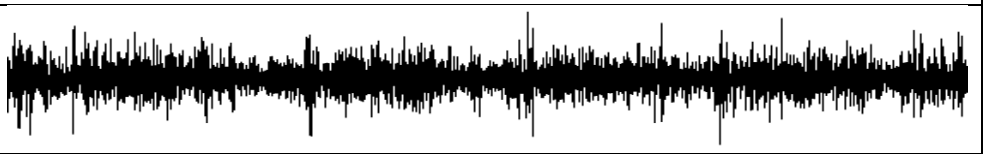
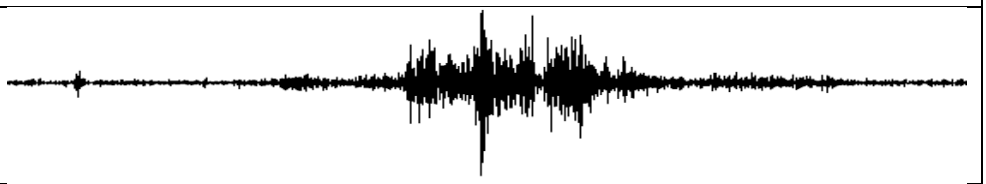
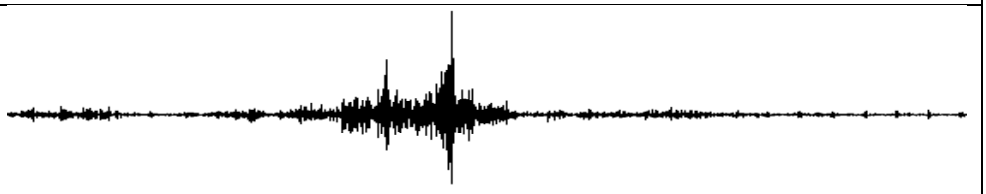


Figure 2.3 EMG electrode placements

Table 2.3 A sample of 30 seconds Chin-EMG activities

Sleep Stages	A Sample of 30 seconds Chin-EMG
W	
N1	
N2	
N3	
R	

2.3 Sleep Stage Classification

The sleep stage classification (SSC) is a process that required trained sleep specialists to divide sleep into stages. Sleep stages provide information about overall sleep quality which can be used in combination with other measures to identify possible sleep disorders. In the SSC, there are a number of selected bio-signals are being determined and evaluated by a sleep specialist. The bio-signals include Electroencephalographic (EEG), Electrocardiogram (ECG/EKG), Electrooculography (EOG), Electromyography (EMG). SSC and the sleep stage scoring can be called interchangeably. The process requires trained sleep specialists to visualise or read the collected bio-signals. It can be very tedious and time consuming process because of the gigantic amount of data collected from a subject during sleep. The size of the data can grow up to more than 2 GB a subject depending on the predefined parameters of each bio-signals.

Since 1968, Rechtschaffen and Kales (R&K) has introduced the first draft of sleep stage scoring guideline. The R&K sleep stage scoring manual was widely used until 2007. R&K divided sleep stages into seven types namely stage 1 (S1), stage 2 (S2), stage 3 (S3), stage 4 (S4), REM, and movement time. R&K was constructed this guideline based on young healthy adults. In addition, the R&K sleep stage scoring manual has been criticised for some unclear points regarding the visual sleep stage classification criterion. It could lead to misclassifications due to misinterpretations by sleep technicians [13] [14]. Later, a refined version of the sleep stage scoring manual was introduced by the AASM. Sleep stages are divided into two groups including Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM). The NREM consists of three sleep stages, NREM-1 (N1), NREM-2 (N2), and NREM-3 (N3). The S3 and S4 in the R&K were combined into N3 in the AASM. There is only one stage in the REM (R), which is equivalent to REM in the R&K manual. Both R&K and AASM scoring manuals measure a sleep stage based on 30-second epochs starting from the beginning until the end of the PSG test [8].

1) Wake Stage (W)

   The wake stage (W) is a sleep stage that a subject still gains consciousness. It can be found in two conditions. Firstly, during eyes-opened, alpha and beta low frequency signals without the rhythmicity of alpha rhythm, mostly in the occipital area. Certain vertical eyes movements can be found. EMG activities are slightly increasing in W compared with other sleep stages. Secondly, during eyes-closed, alpha rhythms are mostly dominant in the occipital area. Slow Eye Movements (SEMs) is presented and the EMG activities are extremely high. Eye blinks are found in W regardless of eye opened or closed.

2) NREM-1 Stage (N1)

   In NREM-1 Stage (N1), the Low Amplitude Mixed-Frequency (LAMF) is presented. For the LAMF, the EEG patterns are ranging from 4-7 Hz. Sleep Spindles and K Complexes are not presented in this stage. For EOG, SEMs is commonly found. During a transition from N1 to N2, Vertex Sharp Wave with Alpha rhythm appears.

3) NREM-2 Stage (N2)

   The K Complexes with arousals are detected in NREM-2 Stage (N2). The durations of both of the Sleep Spindles and the K Complexes are

more than 0.5 sec. Chin EMGs are usually lower than those in W. EOG is inactive. SWA can be visualised. In the AASM recommendations, there are a number of minor criterion that are used to identify N2.

4) NREM-3 Stage (N3)

The characteristics of NREM-3 Stage (N3) are similar to those in N2. However, there are certain conditions that can be used to distinguish N2 and N3. For example, during N3 absences, SWA will last less than 6 sec. A detection of a major body movement followed by SEM ends N3.

5) REM Stage (R)

The REM Stage (R) is a deep sleep stage. The saw-tooth and the alpha rhythm are mostly found in R. Rapid eye movements are also found in R.

Epoch length is a predefined duration for dividing continuous collected bio-signals. According R&K and AASM standards and recommendations, 30 seconds epoch length is recommended and used as a standard divider. However, a number of research works investigated trade-offs of various epoch lengths, processing times, and classification performances. A research mentioned that the 4-second epoch length promoted the best classification of Sleep and Wake [15].

2.4 Data Mining Techniques

Data mining are increasingly integrated into various types of research works including the medicine for a number of years. New data mining techniques are being proposed. However, traditional data mining techniques are still intact. Specific data mining techniques used in our research work will be discussed in more detail. In a broader understanding, there is a recommended scheme for conducting data mining project. In 1996, Cross-Industry Process for Data Mining (CRISP-DM) was firstly introduced and made publicly available as a recommended structured approach for planning a data mining project. The CRISP-DM was supported in European Union Project called the ESPRIT funding initiative. The CRISP-DM has a cycle which consists of six phases [16] [17]. Figure 2.4 shows the CRISP-DM approach.

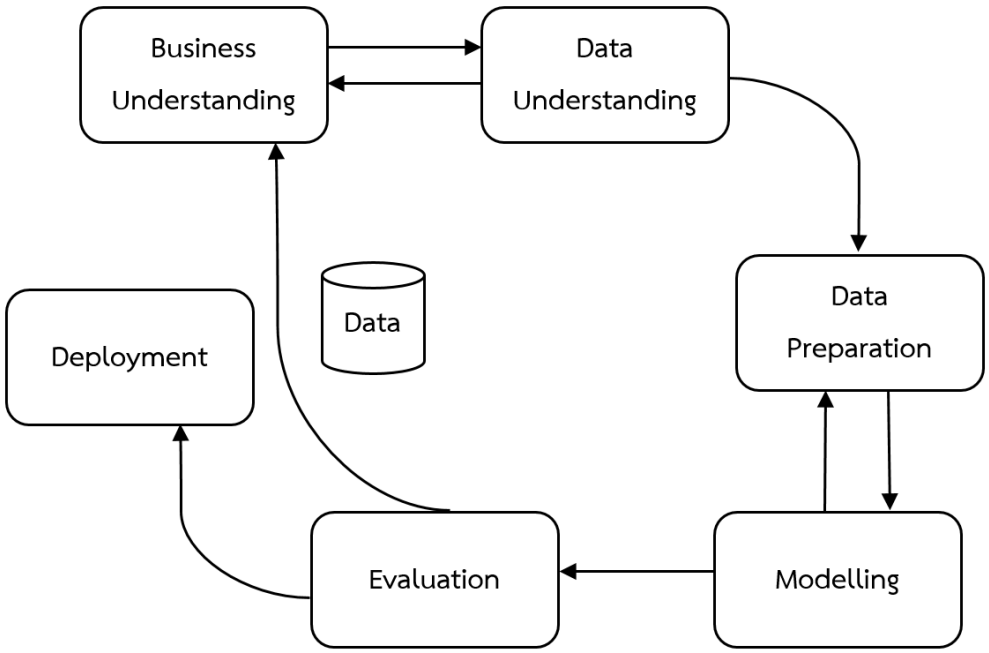


Figure 2.4 CRISP-DM

1) Business understanding

In this stage, primary objectives and goals of a data mining project are defined by accessing current situations of the project. The current situations cover both internal and external dimensions. A set of success criterion are predetermined in order to conduct a project plan.

2) Data understanding

On the completion of data collection, data are studied and explored. Exploratory data analysis is usually performed for capturing the overview of the collected data. The quality of the collected data are evaluated.

3) Data preparation

The collected data are prepared for subsequent processes. The collected data are transformed to proper formats for the data modelling. The data cleansing is usually performed for better data quality. The final processed data are passed to the next phase, modelling.

4) Modelling

In common, several data modelling techniques are selected. The technique is calibrated to gain better performances. If there is no acceptable technique, the process will loop back to the data preparation phase.

5) Evaluation

The data modelling techniques are evaluated based on its performances such as classification performances. Evaluation criterion can be varied depending on the objectives and goals. A decision is made for a selection of a data modelling technique.

6) Deployment

The selected data modelling technique is used to implement the data analysis and show reports. The process can be looped back to the previous ones if the results are not promising. In terms of business related problems, the final results are delivered to customers. On the other hand, for research problems, the final results are used to solve the predetermined objectives and goals.

2.4.1 Decision Tree (DT)

Decision Tree (DT) is a tree-like structure. DT is a supervised learning algorithm. Preclassified target variables must be specified. The target variables include all available subsets in the classification or prediction. A training dataset is also required for establishing a DT model. The training dataset should be varied and covered all specified subsets. In 1984, Breiman, et al developed a new DT model called Classification and Regression Tree (CART). It is a binary classification model. It composes of a root node and two branches in each nodes. A node is divided with an optimal splitting criteria that is constructed from searching all available variables and possible splitting values. The optimal splitting criteria is calculated by the following formula.

$$\phi(s|t) = 2P_L P_R \sum_{j=1}^{\#\ classes} |P(j|t_L) - P(j|t_R)| \qquad (2.1)$$

where

$$t_L = Left\ child\ node\ of\ node\ t$$

$$t_R = Right\ child\ node\ of\ node\ t$$

$$P_L = \frac{Number\ of\ records\ at\ t_L}{Number\ of\ records\ in\ training\ set}$$

$$P_R = \frac{Number\ of\ records\ at\ t_R}{Number\ of\ records\ in\ training\ set}$$

$$P(j|t_L) = \frac{Number\ of\ classes\ j\ records\ at\ t_L}{Number\ of\ records\ at\ t}$$

$$P(j|t_R) = \frac{Number\ of\ classes\ j\ records\ at\ t_R}{Number\ of\ records\ at\ t}$$

Figure 2.5 shows a partial decision tree of a binary classification Oxygen Desaturation (D) and Periodic Limb Movement Disorder (P) from our research work [18]. There are three involving variables including SaO2, Pulse, and O2.A1. In 1984, Quinlan introduced C4.5, an extension of Iterative Dichotomiser (ID3). The C4.5 identifies a node using the same process in the CART. However, there are a number of major improvements that make the C4.5 more superior. The C4.5 is not restricted to the binary classification as presented in the CART. The C4.5 can have more than two branches in one node. Even though the C4.5 performs the same process in identifying a node as in the CART, the C4.5 calculates the optimal split differently.

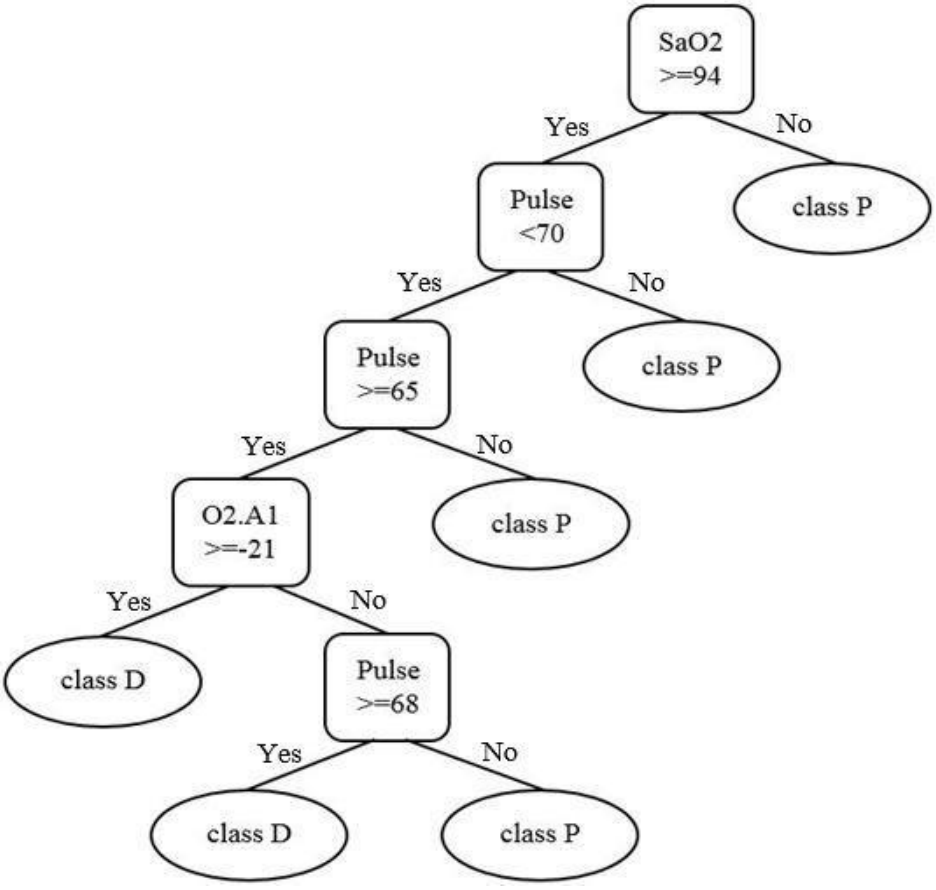


Figure 2.5 A partial decision tree of a binary classification between Oxygen Desaturation (D) and PLMD (P)

The C4.5 utilises information gain or entropy reduction concept to identify the optimal split. The information gain measures “information content” in the studied data partition. Generally, the highest information gain attribute is selected to be the optimal split. The following formula is used to calculate the average amount of information required for classifying a record in *D.* It is also known as the entropy.

$$Info(D) = -\sum_{i=1}^{m} p_i log_2(p_i) \tag{2.2}$$

where

p_i is the probability the D tuple that belongs to class C_i

In order to select a proper attribute, an attribute A is selected for considerations with in the D. The attribute A has v distinct values of $\{a_1, a_2, ..., a_v\}$, as in the training data. If the attribute A contains discrete values, these values harmonise with the v outcomes of a test on the attribute A. Therefore, the attribute A can be utilised as a splitter for dividing D into v distinct groups or subsets, $\{D_1, D_2, ..., D_v\}$. These subsets corresponds to the growing branches from a particular node N. A calculation of $Info_A(D)$ is required. It measures the amount of information required to classify a record with in D based on the partitioning by the attribute A.

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j) \tag{2.3}$$

Finally, the information gain, $Gain(A)$, of the attribute A is measured by the difference between the original information required and the information required after partitioning on the attribute A. An attribute, A, with the highest $Gain(A)$ is selected as a splitter of a particular node N. The followings is the equation for the information gain.

$$Gain(A) = Info(D) - Info_A(D) \tag{2.4}$$

Figure 2.6 shows an ordinary algorithm for constructing a decision tree from a training data. There are three inputs of this algorithm. D is a portion of training data with labelled classes. *Attribute_list* is a set of candidate attributes that one of them can be selected to be the splitting attribute. *Attribute_selection_method* is a procedure to identify a splitting criteria that performs the partitioning of data records into an individual class the best. The splitting criteria consists of a splitting attribute

together with a splitting point or a splitting subset. Initially, a node *N* is created. If all data records in *D* are in the same class *C*, the node *N* is assigned to be a leaf, which is labelled with the class *C* as shown in steps 2 and 3. Terminating conditions, a majority voting, are shown in steps 4 and 5.

Algorithm: Generate_decision_tree. Generate a decision tree from the training tuples of data partition, *D*.

Input:

- Data partition, *D*, which is a set of training tuples and their associated class labels;
- *attribute_list*, the set of candidate attributes;
- *Attribute_selection_method*, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a *splitting_attribute* and, possibly, either a *split-point* or *splitting subset*.

Output: A decision tree.

Method:

```
(1)  create a node N;
(2)  if tuples in D are all of the same class, C, then
(3)      return N as a leaf node labeled with the class C;
(4)  if attribute_list is empty then
(5)      return N as a leaf node labeled with the majority class in D; // majority voting
(6)  apply Attribute_selection_method(D, attribute_list) to find the "best" splitting_criterion;
(7)  label node N with splitting_criterion;
(8)  if splitting_attribute is discrete-valued and
         multiway splits allowed then // not restricted to binary trees
(9)      attribute_list ← attribute_list − splitting_attribute; // remove splitting_attribute
(10) for each outcome j of splitting_criterion
     // partition the tuples and grow subtrees for each partition
(11)     let Dj be the set of data tuples in D satisfying outcome j; // a partition
(12)     if Dj is empty then
(13)         attach a leaf labeled with the majority class in D to node N;
(14)     else attach the node returned by Generate_decision_tree(Dj, attribute_list) to node N;
     endfor
(15) return N;
```

Figure 2.6 Algorithm: decision tree

In step 6, *Attribute_selection_method* determines the splitting criterion and test each of the attributes in *attribute_list* on the data records in *D*. Finally the best splitting criterion, separate the data records in *D* into individual classes, with a selected attribute are identified. The outputs of step 6 are the best splitting criterion, a splitting point or a splitting subset. The best splitting criterion refers to a "pure" partition in which all of the data records in the partition are assigned to the same

class. However, in the real world experiments, a “pure” partition can be found in conjunction with impure or partial pure partitions. Continuously, in step 7, the splitting criterion are used to label the node *N*, which is a test at this node. In steps 8 and 9, if the *splitting_attribute* is discrete values, the tree can be grown in multiple ways, not restricted to the binary tree. The *attribute_list* is adjusted based on the *splitting_attribute*. The rest of the steps show a general concept of branch constructions. A branch is created from a node *N* for each of the outcomes of the splitting criterion. The data records in *D* are separated respectively [19].

A tree can grow indefinitely until satisfactory conditions are met. In many cases, over-growing tree compromises classification performances. Pruning tree is a process of reducing the size of the tree by eliminating weak or less relevant leaves. The objective of the pruning process is to improve overall classification performances and reduce the overfitting problem [16] [17].

2.4.2 *k*-Means Clustering (kMC)

k-Means Clustering (kMC) is an unsupervised learning technique. Unlabeled or ungrouped data are assigned into groups based on specified the number of groups *C*. For instance, *D* is a dataset contains *n* objects. kMC partitions the dataset, *D*, into *k* clusters, C_1, C_2,...,C_{k-1}, C_k. The conditions embeds in the kMC partitioning model are $C_i \subset D$ and $C_i \cap C_j = \emptyset$ for $(1 \leq i, j \leq k)$, It performs in a centroid-based partitioning mechanism. Centroids, the centre of each cluster, are specified.

Algorithm: *k*-means. The *k*-means algorithm for partitioning, where each cluster’s center is represented by the mean value of the objects in the cluster.

Input:

- *k*: the number of clusters,
- *D*: a data set containing *n* objects.

Output: A set of *k* clusters.

Method:

(1) arbitrarily choose *k* objects from *D* as the initial cluster centers;
(2) **repeat**
(3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(4) update the cluster means, that is, calculate the mean value of the objects for each cluster;
(5) **until** no change;

Figure 2.7 Algorithm: k-means

The data are labelled based on feature similarities towards the selected centroid of each clusters. The feature similarities mostly are calculated from distances between data points and each centroids. The closest distance is the indication for the assignment. Figure 2.7 shows an algorithm of k-means [19]. Figure 2.8 shows a classification of randomly generated data sets using kMC with *k*=3.

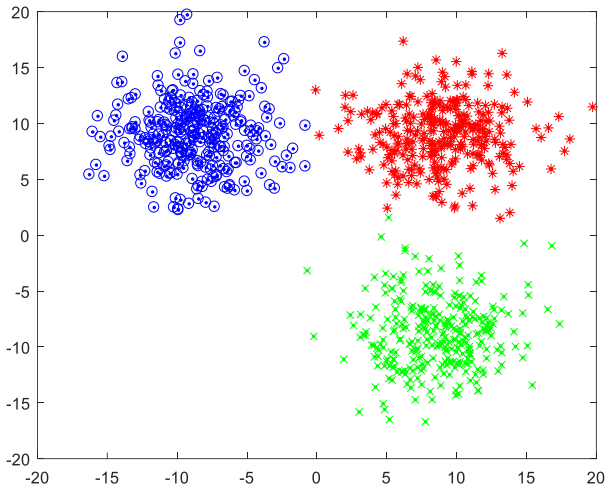


Figure 2.8 A representation of kMC classification with k=3

2.4.3 *k* Nearest Neighbours (*k*NN)

k Nearest Neighbours (*k*NN) is a supervised learning technique. The technique operates in the principle of measuring the distances between neighbours or data points. *k* in the *k*NN is a number of neighbours or data points that are considered to be the same group. The distances between data points are taken into account. One of the mostly used distance measures is the Euclidean Distance. For instance, *a* and *b* are the objects described by n numeric attributes. The Euclidean Distance between *a* and *b* is as follows

Let $a = (x_{a1}, x_{a2}, \ldots, x_{an})$ and $b = (x_{b1}, x_{b2}, \ldots, x_{bn})$

$$dist(a,b) = \sqrt{(x_{a1} - x_{b1})^2 + (x_{a2} - x_{b2})^2 + \ldots + (x_{an} - x_{bn})^2} \quad (2.5)$$

Figure 2.9 shows an algorithm of kNN [20]. A training data set X is specified together with its class labels Y. Unknown samples x are those data point that are required to classify into classes. The variable *k* is a number of nearest neighbour. It must be predetermined and there is no ground rules regarding the *k*. It is likely a trial and error process. Distance measurements between data points are calculated. The

smallest distances are selected based on the number of *k* that is specified. A data point is nominated to a class by a majority vote of its neighbours. The data point is classified to the class most common among its *k* nearest neighbours. kNN has a number of advantages. There is no previous knowledge required for kNN. However, there are a number of disadvantages. A computational cost can be high due to massive calculations of distances between all data points and nearest their neighbours. Finally, one of the most important scenarios that can be found in kNN and clustering technique is the curse of dimensionalities. Specifically, if attributes or features dramatically increase, a number of unrelated features can lead to inefficient classifications.

Algorithm: *k*-Nearest Neighbour. *k*NN is a non-parametric or lazy learning. A data point is classified by a majority vote of its neighbours. The majority vote is based on feature similarity.

Input:

- X: training data
- Y: class labels of X
- x: unknown sample
- *k*: a number of nearest neighbour

Output:

Method:

(1) **for each** instance i of all x
(2) **compute** distance $d(X_i, x)$
(3) **endfor**
(4) **identify** a set of *I* containing indices for the *k* smallest distances $d(X_i, x)$
(5) **return** majority label for $\{Y_i \text{ where } i \in I\}$

Figure 2.9 Algorithm: kNN

2.4.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning technique. Basically, it transforms a training dataset into a higher dimension. It classifies data into groups based on a separating hyperplane. A hyperplane can be both linear and non-linear. There are a number of potential hyperplanes in one particular problem. However, only one of the potential hyperplanes is selected for separating groups. The main selecting condition is the hyperplane that achieves maximum separation. Formally, the SVM algorithm searches for the maximum marginal hyperplane (MMH). The MMH has the largest margins that can form a generalization model. In a binary classification, a hyperplane separate data into two groups in a two dimensional space. In multidimensional spaces, a hyperplane divides data into many groups, which is usually found in real world research works. The MMH expression is shown below [19] [21].

$$\underset{x_i \in D}{argmin}\ d(x) = \underset{x_i \in D}{argmin}\ \frac{|x_i \cdot w + b|}{\sqrt{\sum_{i=1}^{d} w_i^2}} \tag{2.6}$$

D is a training dataset. x_i is a data point with in the *D*. The *w* is a hyperplane direction with an offset scalar *b*. The argmin d(x) determines minimum distances between data points toward a hyperplane. With the minimum distances, it promotes the maximum margin of the hyperplane, which leads to better classification performance. On the other hand, it avoids local minima within a dimensional space. For mathematical calculations, there are three main conditions as follows.

$$\begin{cases} If\ Y_i\ is\ a\ positive\ class; & wx_i + b \geq 1 \\ If\ Y_i\ is\ a\ negative\ class; & wx_i + b < 1 \\ For\ All\ i\ ; & y_i(w_i + b) \geq 1 \end{cases}$$

The SVM kernel function is a mathematical function that plays an important role in SVM and the classification performance. The SVM kernel function transforms data into a specific form. The SVM kernel functions include Radial Basis Function (RBF), Linear, Nonlinear, Polynomial, Sigmoid, etc. The RBF is one of the generally used kernel function. It can be used without prior knowledge about the data therefore the RBF is the firstly most selection kernel function. The RBF can be in various forms including Gaussians, Multiquadric, Inverse quadratic, Inverse multiquadric, etc. The followings is the mathematical expression of the Gaussians RBF. The γ parameter specifies the width of the bell-shaped curve. The γ is usually set as $= \frac{1}{2\sigma^2}$. The σ is generally is an adjustable parameter of the Gaussian RBF. Figure 2.10 shows a classification using SVM. If a training dataset is linearly separable, two parallel hyperplanes, which are used to separate two data classes, are selected. The margin is the area that is bounded by these two hyperplanes.

$$k(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2) \tag{2.7}$$

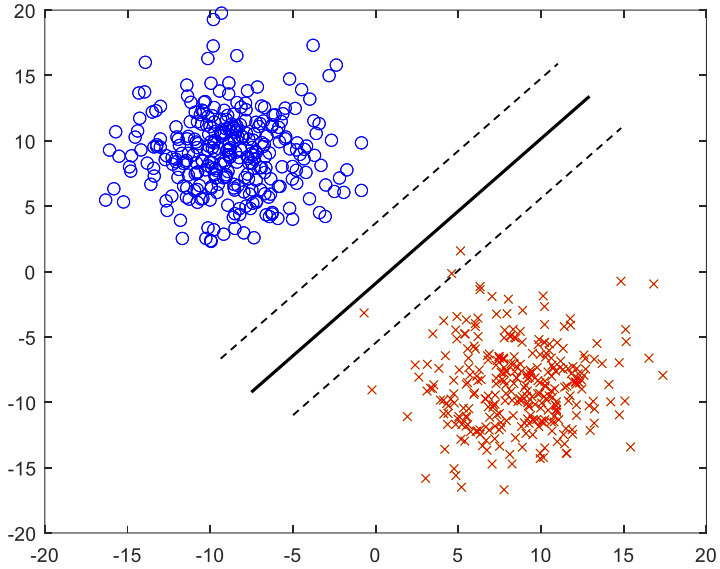


Figure 2.10 A representation of SVM classification

2.4.5 Model Evaluation

This research work evaluates models based on a number of measurements including Accuracy, Sensitivity, Specificity, F-Measure, and Area Under Curve (AUC). AUC has been recommended to use as a main classification measurement especially in the medical or health-related research [19] [22] [23] [24]. The AUC is also suitable for imbalanced datasets, which is a frequently found scenario in real world dataset. All of the selected measurements are designed based on the confusion matrix. There are four basic settings True Positive (TP), True Negative (TN), FP (False Positive), FN (False Negative). In general, TP is a number of positive instances that are classified as positive. FP is a number of negative instances that are classified as positive. TN is a number of negative instances that are classified as negative. FN is a number of positive instances that are classified as negative.

For example, a dataset contains bio-signals from PSG consists of attributes from EEG, EKG, EMG, and EOG. The dataset also includes a target attribute that has two classes, Hypopnea (H) and Non-Hypopnea (Non-H). The objective of this classification is to identify the Hypopnea (H), a condition of partial airway blockage. In this particular dataset, TP is a number of H instances that are classified as H. FP is a number of Non-H instances that are classified as H. TN is a number of Non-H instances that are classified as Non-H. FN is a number of H instances that are classified as Non-H.

The followings are the formulae of the selected performance measures in this research work. Accuracy is a proportion of accurately classified instances. It is

calculated by the combination of TP and TN over overall settings. Sensitivity or Recall or True Positive Rate (TPR) is the ratio of all positives that are correctly identified. Specificity or True Negative Rate (TNR) is all of the negatives that are correctly identified as negatives. F Measure is the harmonic average of the precision and the recall of the test case. The best F Measure score is 1, the best precision and recall, and the worst F Measure is 0. AUC is calculated based on a combination of Sensitivity and Specificity over 2 [19].

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{2.8}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{2.9}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{2.10}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2.11}$$

$$F\ Measure = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} \tag{2.12}$$

$$Area\ Under\ the\ Curve\ (AUC) = \frac{Sensitivity + Specificity}{2} \tag{2.13}$$

Apart from the above measurement, a model evaluation requires a formal methodology. The *k*-fold cross validation method is one of the mostly selected method in classification and prediction problems. Practically, a dataset is randomly divided into k portions (P_1, P_2, ..., P_k). In each fold, a portion (P_i) is reserved as a test set. Explicitly, all data portion (P_1, P_2, ..., P_k) are tested exactly once and they becomes a part of the training set k-1 times. The k-fold cross validation method is an iterative process. It operates until all of the portions has been tested. The training sets in each fold are employed to build a classification model. The test sets are utilised to evaluate the performance of the classification model using proper measurements [19]. Figure 2.11 is a representation of 10-folds cross validation method.

Figure 2.11 A representation of 10-folds cross validation method

2.5 Literature Review

In the PSG test, a combination of bio-signals or features are collected for both classifying sleep stages and diagnosing sleep-related symptoms. This research work proposes two models that are ASSC and ASDC. There were various data preprocessing in research works. Some only selected partial features for their studies. On the other hand, full features from the PSG test were used. Some techniques are being used in the data preprocessing step such as wavelet transform, spectral entropy, and other mathematical and statistical techniques. Some features including mean, median, mode, kurtosis, standard deviation are also widely included in the data preprocessing step. Classification models in both ASSC and ASDC can be constructed using one model or multiple models.

The ASSC has been comprehensively studied in various research works. The ASSC, sometimes called Computer Aided Sleep Stage Classification, leads to correctly classify sleep disorders. In general, ASSCs proposed methods to classifying sleep stages automatically. In a standard SSC, errors can be found in during the sleep stage classification or scoring process. Therefore, there are a number of research works that presented a comparison of different sleep specialists that read on bio-signals from same patients. For example, a dataset from ISRUC-Sleep database, which was collected from the Sleep Medicine Centre at the Hospital of Coimbra University, consists of 100 subjects with two corresponding sleep stage classification results. The results were evaluated by two trained sleep specialists [25]. This particular datasets leads to many research works that attempted to find the correlations between the two trained sleep specialists. It has the final goal of minimising human errors. Another

research work proposed a hybrid approach, SVM-DT. The approach combined SVM into DT as a DT prepruning method. The researchers believed that this approach would minimise the scoring errors [26]. In addition, the multi-class SVM was highly used in various research works. A research work introduced a Dendrogram based SVM (DSVM) framework. The DSVM utilised the multi-class SVM classification based on a DT approach. A number of features were selected. The framework was evaluated with the k-fold cross-validation. The result was promising if it was compared with manual scoring techniques [27]. Another research work utilised the EEG signals to classify sleep stages using the multi-class SVM to classify sleep stages from EEG signals. It gained 92% accuracy [28]. Additionally, Sirvan Supervised Method for Sleep Staging (SSM4S) was published. It utilised the multi-class SVM with selected features from EEG, EMG and EOG. The SSM4S combined two SVM models for classifying sleep–wake stages and classifying multi-class sleep stages [29].

ASDC research works are also widely conducted and investigated. Some research works attempted to use minimum or selective bio-signals for sleep disorder classifications. In 1999, an investigation of the correlations between heart rates, oxygen desaturation levels and the apnoea durations in infant. This research work was studying 236 apnoea epochs and found that both heart rates and oxygen desaturation levels were significantly related to the apnoea durations [30]. This research work was compromised by other works due to a very small sample of subjects. Moreover, the work is based on subject-independent study, which could be difficult to implement in later clinical studies. Another research work was performed to identify OSA with the 24 hours ambulatory ECG. The ambulatory ECG, a limited capability version of the full ECG. Time and frequency domains of Heart Rate Variability (HRV) analysis techniques were used. The research work was conducted on 95 subjects, 48 with OSA and 47 without OSA. The results were 81.25% sensitivity, 46.81% specificity and 64.21% positive predictive value [31]. This research work was designed for a very specific medical equipment, the ambulatory ECG. This could compromise the classification result. However, it could benefited hospitals that were not equipped with PSG. A group of selected features were used with a classification process using MLP and flexible decision rules. The result showed 82% accuracy in the deep and paradoxical sleep [32]. Another similar research setting employed the Neural Network (NN) system to validate the oxygen desaturation levels from oximeters for OSA predictions. The overall classification result was 93.30% accuracy [33].

The emergence of smartphones leads to a new era of medical research related works. Researchers use smartphones as a tool for analysing certain symptoms or sending data from other devices to smartphones for storage. SleepAp, a mobile

device-based OSA screening test was introduced. A group of researchers developed a mobile application that records audio, actigraphy and Photoplethysmogram (PPG) signals via a mobile oximeter. The mobile oximeter connected to a mobile phone via the Bluetooth connection. PPG can be used to detect blood volume changes by measuring skin illuminations. The PPG is the simplest and a cost effective method. The classification using SVM reached 88.40% accuracy in the first experiment and 92.30% accuracy in the second experiment [34]. Another of our previous research work also using a smartphone to detect the OSA during sleep by analysing the snoring sound. The *kMC* was solely used in this research work. The number of *k* clusters was set based on the number of sleep disorder classes in each sleep stages in our experiment [35]. Table 2.4 shows a comparison of ASSC and ASDC research works.

Table 2.4 A comparison of ASSC and ASDC research works

Research work by	Techniques and Models	Strengths	Weaknesses
Automatic Sleep Stage Classification			
Khalighi, et al [25]	● ISRUC-Sleep database is publicly provided. The database contains various datasets. The publication is related to the announcement of ISRUC-Sleep database.	● 100 subjects. ● A comparison of two sleep specialists sleep stage scoring. ● Make it available publicly.	● There is no classification or predictive models available.
Lajnefa, et al [27]	● ASSC with a multiclass SVM classification based on a decision tree approach.	● Both time and frequency domain features were used. ● Feature selection: forward sequential selection. ● k-fold cross-validation.	● Only 15 subjects. ● Only subject-independent classification.
Aboalayon and Faezipour [28]	● ASSC with multiclass SVM	● Only EEG signal. ● 92% accuracy.	● Only 13 subjects. ● Only subject-independent classification. ● If other signals

			apart from EEG are included, it may improve the classification performance.
Khalighi, et al [29]	• ASSC using SVM with a comparison with LDA, Naïve Bayes (NB), and AdaBoost. • Two main classification models: sleep–wake detection and multiclass sleep staging.	• Only 6 channels for sleep-wake detection. • Only 9 channels for multiclass sleep staging.	• Only subject-independent classification (Global performance).
Chapotot and Becq [32]	• ASSC using artificial neural network classification and flexible decision rules.	• An implementation of majority vote among ten consecutive classifiers. • Only EEG and EMG. • 82% accuracy.	• Only 48 subjects. • 20s epoch length, which is not based on ASSM recommendation. • Only subject-independent classification. • Unknown classifier selection process.
Automatic Sleep Disorder Classification			
Behar, et al [34]	• ASDC using SVM	• A mobile phone application as a means to collect bio-signals during sleep. • Large dataset: 735 subjects both healthy and non-healthy (with OSA).	• Only classifying OSA. • Only subject-independent classification. • Patients need to wear certain medical equipment during sleep at home.
Wongsirichot, et al [35]	• ASDC using kMC.	• A mobile phone application was designed to analyse snoring sounds in order to	• Small dataset. • Only subject-independent classification. • Not very high

		detect OSA and Hypopnea. ● It was developed as a sleep disorder screening tool.	accuracy due to various reasons including natural noise during sound collection and some low quality built-in microphones of mobile phones.
Wongsirichot and Hanskunatai [18]	● Investigation studies of ASDC.	● Classification performance comparisons of kMC, SVM and CART were conducted.	● No feature selection. ● Only subject-independent classification.

Chapter 3
Research methodology

This research work proposes two main classification models, ASSC and ASDC. The objective of this research work is to investigate existing machine learning techniques and develop a new hybrid machine learning technique for improving the overall process of ASSC and ASDC. The ASDC research works investigate various combinations of ML techniques and specific epoch length. Moreover, the assurance of improved ASSC promotes overall sleep disorder classification subsequently. The research work is designed based on the CRISP-DM methodology.

3.1 Business understanding

An ordinary Sleep Stage Classification (SSC) and Sleep Disorder Classification (SDC) process is tedious and time-consuming. A trained sleep specialist is required for interpreting collected bio-signals during a PSG test. Automatic Sleep Stage Classification (ASSC) or Computer-Aided Sleep Stage Classification becomes an interesting topic in this area. With an ASSC, it possibly speeds up the process of sleep stage classification. An ASSC usually comprises of mathematical, statistical and computational models. For the computational models, selected models mostly are based on data mining techniques and related computer science practices. For example, a research work related to ASSC was using SVM and Decision Tree [26]. One of the main objectives of the ASSC is to reduce the processing time and achieve quantitatively comparable or superior results. On the other hand, the Sleep Disorder Classification is another popular topic in this field. Researchers attempted to find automatic analytical models to identify various types of sleep disorders including OSA. Commonly, the models are derived from various mathematical and statistical models and machine learning techniques.

In our research works, we propose a newly developed ASSC using a special Hybrid Machine Learning Model (HMLM). Moreover, we also propose in-depth investigations in ASDC. Various well-known ML techniques are selected and studied. In both ASSC and ASDC, the feature selection and extraction plays an important role. Features with high impact towards the classification results are identified. Specific epoch lengths are also considered. In addition, the model evaluations are performed based on two contexts, the subject-independent and the subject-dependent classifications. The subject-independent considers the overall classification performances of studied classification models regardless of subjects. On the other

hand, the subject-dependent looks into specific subjects. Specifically, it considers a classification model performing on a particular subject rather than all subjects.

3.2 Data understanding

Both of our proposed ASSC and ASDC models are conducted with PSG datasets. PSG datasets was collected from sleep laboratories. The data are collected from real subjects. The patient confidentiality and the Privacy Acts are applied. However, a number of PSG datasets are publicly available. For the ASSC experiment, the ISRUC-Sleep Subgroup I dataset was selected [36]. It contained 8-9 hours of bio-signals from 100 subjects. The subjects comprised of 53 males, 42 females and 5 unspecified sex. The subjects included both healthy and OSA-related. Demographic data of the subjects were age 51±16 years, BMI 23.53±12.83 kg/m^2, height 1.35±6.63 m, and weight 65.09±34.97 kg. Technically, the ISRUC-Sleep data was sampled at 200 MHz and divided into 30s epoch length. The collected bio-signals included EEG, EOG and EMG as shown in Table 3.1. In addition, the ISRUC-Sleep Subgroup I dataset was collected from active subjects. It shows degrees of imbalance. Table 3.2 shows a number of instances categorised by sleep stages. Additionally, there are five classes in the AASC problems, W, N1, N2, N3 and R, respectively.

For the ASDC experiment, the PSG dataset was collected from the Songklanagarin Hospital, Hat Yai, Songkhla, Thailand. The dataset was recorded via the Sleepscan VISION software [37]. It contained full night recordings of five subjects, two males and three females. Demographic data of the subjects are age 39.00±11.20 years, BMI 27.75±4.50 kg/m^2 and AHI 8.20±2.00. There are 440,593 records. The records were divided into 30s epoch length. There were four classes of sleep disorders classified by a sleep specialist in this dataset, including Oxygen Desaturation (D), Hypopnea (H), Isolated Limb Movement (I) and Periodic Limb Movement (P). It was also scored into sleep stages, W, N1, N2, N3, and R. The attributes in the dataset are shown in Table 3.3.

Table 3.1 ISRUC-Sleep Subgroup I Bio-Signals

Bio-Signal	Description
Electroencephalography (EEG)	
C3-A2	Monopolar EEG in a position C3-A2
C4-A1	Monopolar EEG in a position C4-A1
F3-A2	Monopolar EEG in a position F3-A2
F4-A1	Monopolar EEG in a position F4-A1
O1-A2	Monopolar EEG in a position O1-A2
O2-A1	Monopolar EEG in a position O2-A1
Electrooculography (EOG)	
LOC-A2	Left Outer Canthus
ROC-A1	Right Outer Canthus
Electromyography (EMG)	
Chin EMG	Placed between the chin and lower lip
Leg-1 EMG	Left leg movement
Leg-2 EMG	Right leg movement

Table 3.2 A number of instances categorised by sleep stages

Sleep Stage	No. of Instances	Percent
W	18418	20.11%
N1	11322	21.12%
N2	28002	12.99%
N3	17532	32.12%
R	11913	13.66%
Total	87187	100.00%

Table 3.3 Songklanagarin Hospital Dataset

Bio-Signal	Description
Electroencephalography (EEG)	
C3-A2	Monopolar EEG in a position C3-A2
C4-A1	Monopolar EEG in a position C4-A1
F3-A2	Monopolar EEG in a position F3-A2
F4-A1	Monopolar EEG in a position F4-A1
O1-A2	Monopolar EEG in a position O1-A2
O2-A1	Monopolar EEG in a position O2-A1
Electrooculography (EOG)	
LOC-A2	Left Outer Canthus
ROC-A1	Right Outer Canthus
Electromyography (EMG)	
Chin EMG	Placed between the chin and lower lip
LAT / RAT	Left and Right anterior tibialis
CHEST	Chest movement
ABDO	Abdomen movement
Electrocardiography (ECG)	
ECG	Standard electrocardiogram signals
Pulse	Pulse
Thoracic respiratory efforts	
CANR	Nasal airflow
FLOW	Mouth airflow
SAO2	Oxygen desaturation
SNR	Amplitude of snoring sounds

A number of features in the Songklanagarin Hospital are more than the ISRUC-Sleep due to different PSG settings. The Thoracic respiratory efforts were another group of signals that were collected from the PSG at the Songklanagarin Hospital. The Thoracic respiratory efforts allow sleep specialist to investigate various airflow patterns including snore amplitude, nasal and mouth airflows and Oxygen desaturation.

3.3 Data preparation

Proposed ASSC and ASDC have a number of slightly different data preparation processes. For the ASSC, there are three main steps in the data preparation step as shown in Figure 3.1.

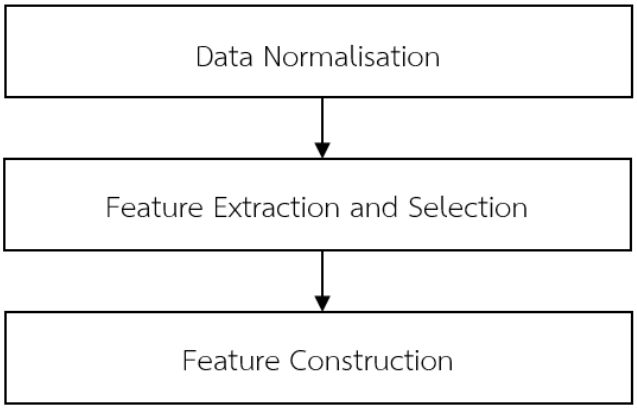


Figure 3.1 ASSC data preparation

1) Data Normalisation

All of the collected bio-signals are based on 6000 continuous values within 30 seconds epoch length. However, each of the bio-signals has different scales. There is a possibility that some of the bio-signals can dominate others, which can lead to bias in the classification model. The data normalisation is used to scale all of the bio-signals between 0 and 1 using the following formula.

$$X' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{3.1}$$

2) Feature Extraction and Selection

Due to the huge amount of collected data per subject, feature extraction and selection is compulsory in ASSC. It assists to select an important features, which have outstanding characteristics towards the classification. Specifically, the feature extraction of a PSG dataset can be considered in two domains, frequencies and amplitudes. Researchers proposed and attempted to use both single and multiple feature extraction techniques however the classification results are still inconclusive [25]. In this research work, the Fast Fourier Transform (FFT) is selected together with the Maximum Amplitude Analysis (MAA) for the feature extraction. All of the PSG attributes were selected. The FFT transform the original continuous signal from the time domain to the frequency domain [19]. The FFT was conducted in the MATLAB that aligns with the original Fourier Transform technique. FFT has been used in various research works in the past with promising results [38] [39] [29] [40]

[36] [41] [42]. On the completion of FFT, the MAA was performed. The MAA is basically located the maximum amplitude in a specific epoch length. In our proposed ASSC model, the epoch length was divided into four different scales including 5, 10, 15 and 30s. Figure 3.2 shows the maximum amplitude of C4-A1 within two consecutive 30 seconds epoch. Table 3.4 shows a sample of N3 Signal in all of the studied epoch lengths.

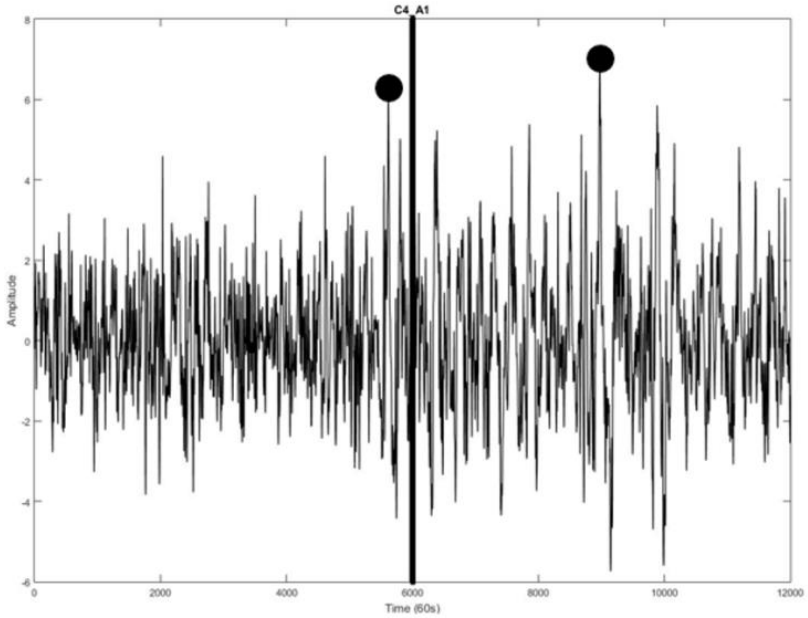


Figure 3.2 Maximum amplitude of C4-A1 within two consecutive 30 seconds epoch

Table 3.4 A Sample of N3 Signal in all of the studied epoch lengths

Ep (s)	C3–A2	C4–A1	F3–A2	F4–A1	O1–A2	O2–A1	LOC–A2	LOC–A1	Chin EMG	Leg-1 EMG	Leg-2 EMG	*SS*	*SW*	*NR*
30	0.913	0.966	0.988	0.972	0.977	0.998	0.980	0.971	0.985	0.951	0.987	*N3*	*S*	*N*
15	0.575	0.515	0.511	0.513	0.563	0.511	0.520	0.509	0.496	0.508	0.510	*N3*	*S*	*N*
	0.913	0.966	0.988	0.972	0.977	0.998	0.980	0.965	0.985	0.951	0.987			
	0.575	0.515	0.511	0.513	0.563	0.511	0.518	0.509	0.496	0.508	0.510			
10	0.885	0.966	0.965	0.956	0.977	0.983	0.980	0.965	0.985	0.951	0.987	*N3*	*S*	*N*
	0.913	0.958	0.988	0.972	0.972	0.998	0.968	0.971	0.984	0.951	0.987			
	0.557	0.515	0.511	0.513	0.563	0.480	0.518	0.509	0.457	0.485	0.510			
	0.575	0.494	0.491	0.486	0.503	0.511	0.494	0.507	0.496	0.508	0.480			
5	0.557	0.484	0.511	0.484	0.487	0.468	0.520	0.507	0.478	0.488	0.500	*N3*	*S*	*N*
	0.885	0.966	0.965	0.956	0.977	0.983	0.980	0.965	0.985	0.951	0.987			
	0.899	0.915	0.988	0.945	0.887	0.998	0.920	0.889	0.924	0.951	0.933			
	0.913	0.958	0.968	0.972	0.972	0.978	0.968	0.971	0.984	0.947	0.987			

3) Feature Construction

A new feature construction is designed and implemented in this research work. Two additional features are added, *SW* and *NR*. The *SW* is

built to classify Sleep and Wake. The *SW* comprises of two values, Sleep(*S*) and Wake(*W*). The *NR* is constructed in order to identify NREM and REM. The *NR* consists of two values, NREM(*N*) and REM(*R*). The following shows the feature construction model. Table 3.5 shows a sample of the dataset after the feature construction process.

$$Let\ \mathbb{U}\ be\ a\ series\ of\ sleep\ stages\ (W, N1, N2, N3, R):$$

$$\mathbb{U} = \{W, N1, N2, N3, R\}$$
$$NREM = \{N1, N2, N3\}$$
$$REM = \{R\}$$
$$WAKE = \{W\}$$

$$Thus, \quad NREM \subseteq \mathbb{U}, REM \subseteq \mathbb{U}, WAKE \subseteq \mathbb{U}$$

$$Let\ x\ be\ a\ instance\ of\ \mathbb{U}, such\ that$$

$$SW = \begin{cases} S, & when\ x \in (NREM \cup REM) \\ W, & when\ x \in WAKE \end{cases}$$

$$NR = \begin{cases} N, & when\ x \in NREM \\ R, & when\ x \in REM \end{cases}$$

Table 3.5 A sample of the dataset after the completion of feature extraction

C3-A2	C4-A1	...	Sleep Stage	SW	NR
0.025996	0.050164	...	N3	S	NREM
0.045462	0.027926	...	R	S	REM
0.024104	0.048959	...	N3	S	NREM
0.017825	0.013538	...	W	W	-

For the ASDC, the main objective is to identify available sleep disorders in the dataset. Moreover, sleep disorder patterns in each sleep stages are investigated. There are three main steps in the data preparation step as shown in Figure 3.3.

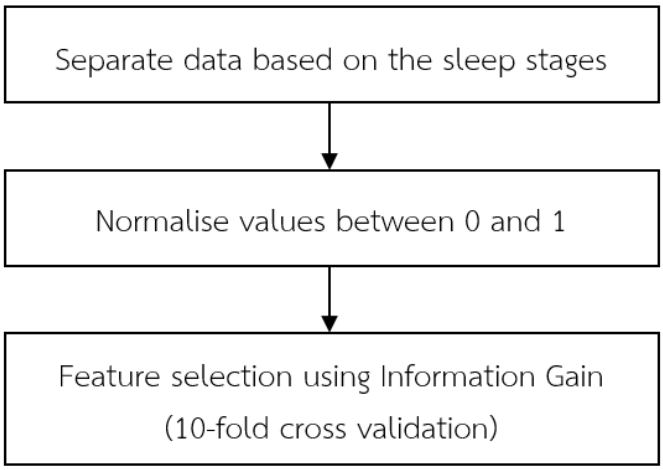


Figure 3.3 ASDC data preparation

Firstly, collected bio-signal data was grouped by sleep stages, N-1, N-2, N-3, and R. Each data record consisted of discrete signal values in specific time series. Figure 3.4 shows sampled consecutive data records within a timestamp.

Time-stamp	C3-A2	C4-A1	…	Sleep disorder
10478.75	-0.1874	7.4874	…	D
10478.75	-1.7853	7.7363	…	D
10478.75	-3.7690	7.3395	…	D
…	…	…	…	…
16032.01	-18.449	0.9919	…	P
16032.01	-17.258	6.9429	…	P
16032.01	-17.227	7.2478	…	P

Figure 3.4 Sampled consecutive data records

The original data were in different scales. The second step of data preparation was designed. The concept of data normalisation was performed to rescale all values between 0 and 1. The data normalisation formula is shown in equation 3.1. Lastly, a feature selection process was performed. The Information Gain (IG) was selected as the feature selector. The Information Gain (IG) is a heuristic function that is used to measure the abilities of attributes or features in classifying data. The higher IG features are better attributes or features for that particular classification problem. The IG formula is shown in equation 2.4 [19] [43]. IG was calculated to measure each of the features with the 10-fold cross validation method for each of the sleep stages. Therefore, each of the feature has ten IGs from the calculations. The average IG of each of the feature were calculated and used to rank all of the attributes. Table 3.6 shows average IG of each feature in sleep stages.

Table 3.6 Average IG of each features in sleep stages

Features	Information gain				
	N-1	N-2	N-3	R	Avg
PULSE	0.297	0.300	0.066	0.675	0.335
SAO2	0.197	0.192	0.087	0.594	0.268
CANR	0.050	0.042	0.015	0.402	0.127
CHEST	0.043	0.040	0.018	0.266	0.092
CHIN	0.030	0.004	0.016	0.272	0.081
ABDO	0.062	0.052	0.020	0.156	0.073
ROC	0.170	0.027	0.037	0.054	0.072
FLOW	0.040	0.030	0.016	0.142	0.057
F3-A2	0. 150	0.018	0.008	0.015	0.048
ECG	0.140	0.010	0.002	0.020	0.043
O1-A2	0.029	0.043	0.030	0.044	0.037
LOC	0.025	0.027	0.031	0.047	0.033
O2-A1	0.006	0.038	0.025	0.055	0.031
C3-A2	0.017	0.034	0.017	0.013	0.020
C4-A1	0.007	0.011	0.010	0.015	0.014
LAT	0.005	0.011	0.004	0.026	0.012
F4-A1	0.007	0.007	0.010	0.019	0.011
RAT	0.004	0.007	0.001	0.008	0.005
SNR	0.003	0.006	0.000	0.000	0.002

According to Table 3.6, in order to select proper numbers of features for classification models. A number of rounds of calculations are designed. In the first calculation round, all of the 19 features are used. In the second calculation round, the lowest averaged IG feature was removed, SNR. Therefore, only 18 selected features remains in the calculation. The iterative calculation continues until only last two highest IG features, Pulse and SAO2, remains in the calculation. Figure 3.5 represents the overall classification accuracy results with selected features in different ML techniques. The graphs can be classified into three common patterns. Firstly, the number of features increase makes the accuracies higher. Secondly, the accuracies dramatically deduct when the number of features are added. Finally, there is no correlation between the changes of the number of features and the accuracies. The kNN classification results reveal that the increases of the number of features reflect the accuracies. The kNN becomes the winner classifier in this experiment. On the opposite side, the kMC classification results reduce when the number of features expand, which is in the second case.

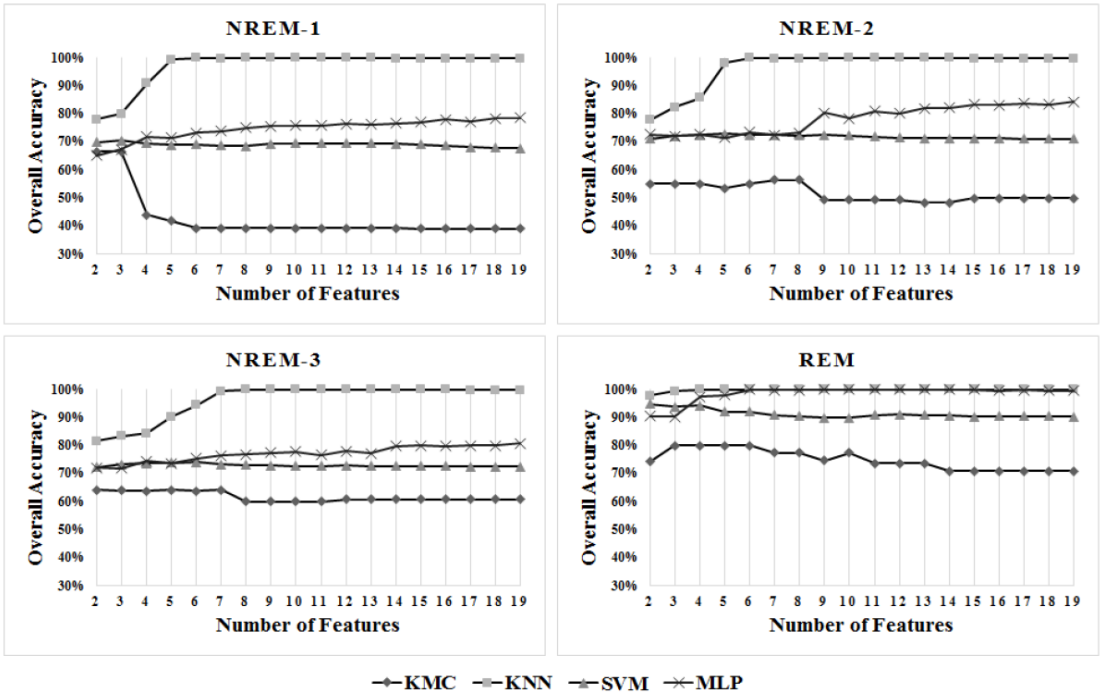


Figure 3.5 Preliminary classification results with selected features

Table 3.6 shows the first two highest average IG features were PULSE and SAO2 (Avg-IG = 0.335 and 0.268). The following features obtained the average IGs less than 0.13. By all means, PULSE and SAO2, are chosen as the nominated features. In terms of medical devices required, PULSE and SAO2 can be gathered by an ordinary oximeter [14]. This can be an advantage for future research works in terms of acquiring expensive medical devices. The following features after PULSE and SAO2 based on the averaged IG ranking are CANR, CHEST and CHIN, respectively. All of the followed features present relatively similar averaged IGs. However, CANR and CHEST are collected by an ordinary ECG and CHIN is monitored by the EMG sensor, which firmly attached to a patient's chin. The CHIN promotes less patient comfortability. With the considerations of minimum uses of medical equipment and patient comfortability, PULSE, SAO2, CANR and CHEST are selected as the optimal features. The number of selected features are used together with kNN. Table 3.7 shows four sets of features in our experiment.

Table 3.7 A number of selected features in proposed ASDC models

Name	Number of features	Features	Description
Minimum *k*NN(*2f*)	2	PULSE, SAO2	Minimum use of medical equipment.
Full Optimal *k*NN(*5f*)	5	PULSE, SAO2, CANR, CHEST, CHIN	First five ranked features based on averaged IG
Optimal *opt-k*NN	4	PULSE, SAO2, CANR, CHEST	Remove CHIN in order to promote patient comfortability
Best *best-k*NN	max	-	Use maximum number of features in order to achieve highest classification results.

3.4 Modelling

For the ASSC, the motivation of our research work is to combine the sophisticated multiple ML techniques with the existing manual SSC process taken by a sleep technician. The selected multiple ML techniques are acting as a hybrid ML model. In addition, a designed Multi-Layer Hybrid Machine Learning Model (MLHM) are designed to accommodate the SSC process. The MLHM is shown in Figure 3.6.

Five sleep stages, W, R, N1, N2 and N3 are included in the current AASM scoring recommendations. In the manual SSC, a sleep specialist initially scores sleep stages into three main groups, sleep–wake, NREM–REM and N1–N2–N3 [8] [11]. In order to follow the manual SSC, two additional attributes, *SW* and *NR*, were constructed as explained in the Data Preparation: Feature Construction. SW and NR are used in the Hybrid Classification Model resided in the MLHM. The *SW* is used in the Layer 1 in the MLHM. A model, which will be discussed later, classifies the Sleep (S) and Wake (W). The output for Layer 1 is W. All data that were classified as S will be passed to the next layer. The data that were classified as S ideally consist of two types, NREM and REM. Layer 2, the *NR* attribute is used to classify the NREM and REM. Only the data that has been selected as S are considered to classify in the Layer 2. The REM is the output of the Layer 2. All of the data that are marked as NREM will be passed to the next layer. Layer 3 classifies the data from Layer 2 into N1, N2 and N3.

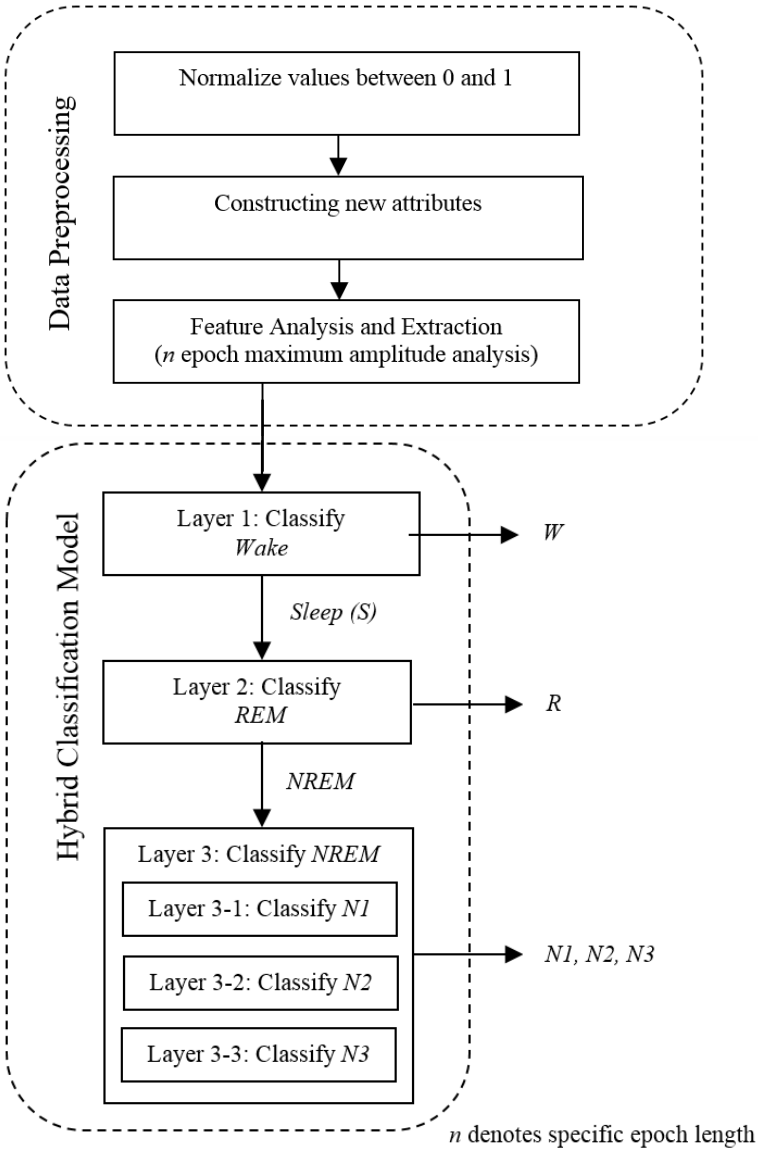


Figure 3.6 Multi-Layer Hybrid ML Model (MLHM)

Each of the Layer in the Hybrid Classification Model in the MLHM has been thoroughly designed. Two ML techniques that are selected for this study are DT and SVM. The main concern is to identify the appropriate epoch length and the selected ML technique. The dataset is divided into 70:30. The first 70% are used as the training set and the rest unseen 30% is used as the test set. Four different epoch lengths, 5, 10, 15 and 30s, are tested with DT and SVM, respectively. The AUC is used as the measurement for selecting the best classifier. Table 3.8 shows the best classifier together with the epoch length in each of the layer.

Table 3.8 The best classifier with the epoch length in each of the layer

Layer	ML(epoch)	ACC	Sens	Spec	*F*	AUC
*L*1	DT(5ep)*	0.633	0.7688	0.2286	0.7026	0.7701
(*W*)	DT(10ep)	0.628	0.7639	0.2352	0.6997	0.7644
	DT(30ep)	0.626	0.7577	0.2400	0.6937	0.7589
	SVM(5ep)	0.444	0.4647	0.5367	0.4303	0.4640
	SVM(30ep)	0.295	0.1800	0.8200	0.1494	0.1800
*L*2	SVM(5ep)*	0.793	0.9038	0.0951	0.8394	0.9044
(*R*)	DT(5ep)	0.764	0.8695	0.1084	0.8184	0.8806
	DT(10ep)	0.748	0.8465	0.1300	0.8078	0.8583
	DT(15ep)	0.745	0.8418	0.1394	0.8024	0.8512
	DT(30ep)	0.740	0.8349	0.1459	0.7972	0.8445
*L*3	DT(15ep)*	0.681	0.8124	0.1866	0.7719	0.8124
(*N*1, *N*2, *N*3)	DT(5ep)	0.680	0.8101	0.1917	0.7708	0.8100
	DT(30ep)	0.676	0.8032	0.1981	0.7675	0.8031
	DT(10ep)	0.676	0.7978	0.2032	0.7729	0.7977
	SVM(5ep)	0.656	0.7982	0.2038	0.5313	0.7981

DT(5ep), DT with 5 sec epoch length, achieves the highest AUC of 0.7701 in Layer 1. SVM(5ep), SVM with 5 sec epoch length, gets the highest AUC of 0.9004 in Layer 2. In Layer 3, DT(15ep), DT with 15 sec epoch length achieves the highest AUC of 0.8124. Therefore, DT(5ep), SVM(5ep) and DT(15ep) are placed in the Layer 1, 2 and 3, respectively. The formation of the Hybrid Classification Model in the MLHM is shown in Figure 3.7.

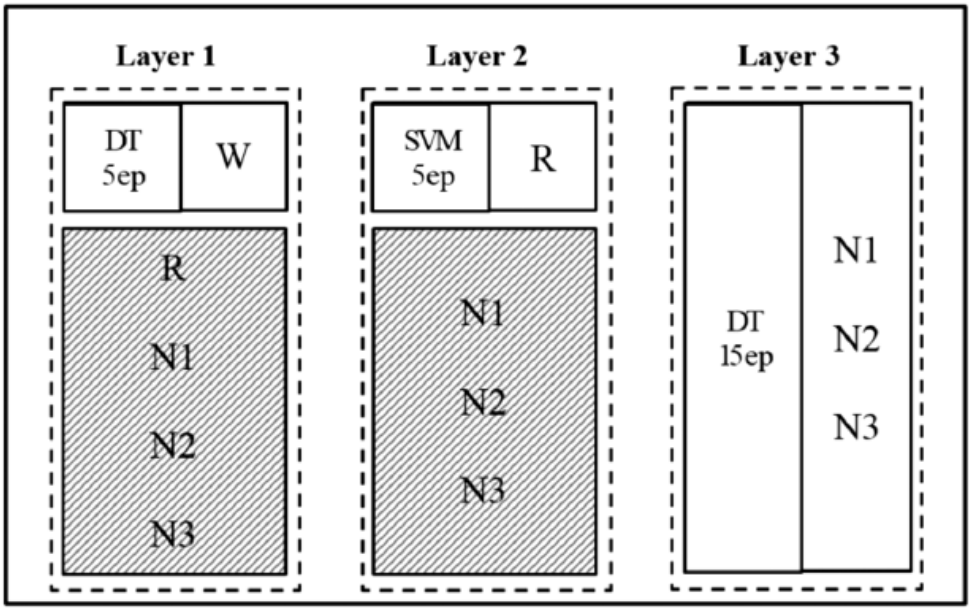


Figure 3.7 Hybrid classification model in MLHM

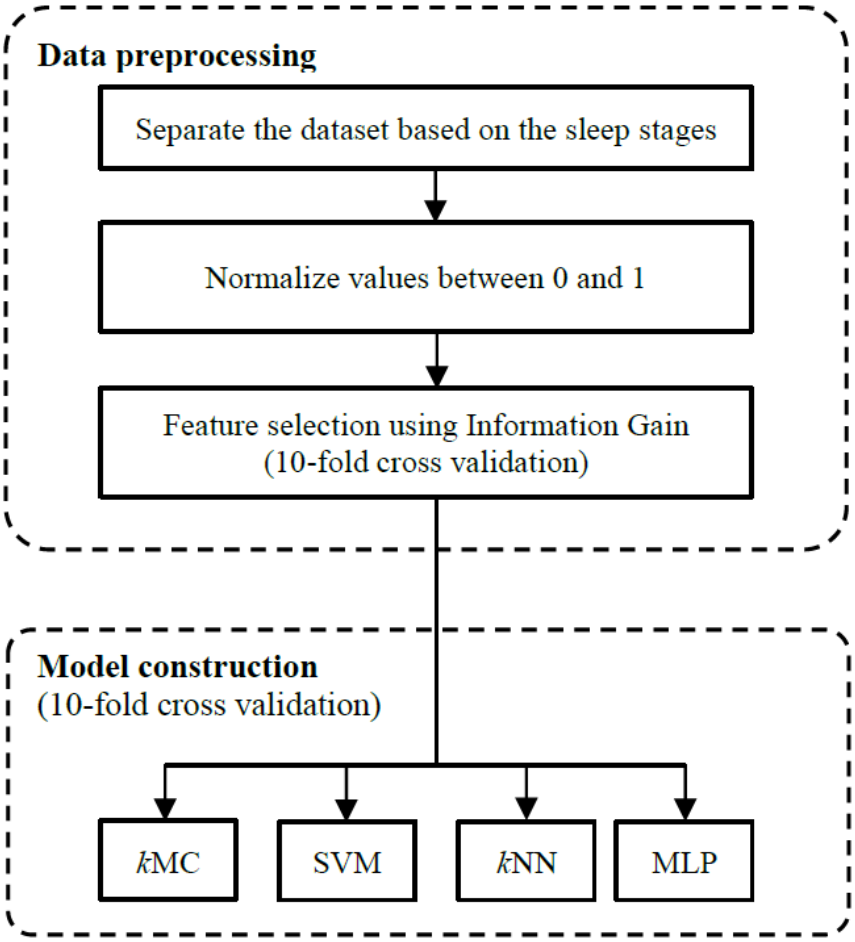


Figure 3.8 ASDC experiment model

For the ASDC, the objective is to automatically identify various types of sleep disorders including Oxygen Desaturation (D), Hypopnea (H), Isolated Limb Movement (I) and Periodic Limb Movement (P) as available in the selected dataset. The proposed ASDC investigates all of the mentioned sleep disorders in all sleep stages using kMC, SVM, *k*NN, and MLP. A reason for selecting these well-known ML techniques is to study the sleep disorder classification performances. This study will potentially explore into more complex methods. The ASDC experiment model is shown in Figure 3.8.

3.5 Evaluation

Both ASSC and ASDC are evaluated based on non-biased methods. In general, the original dataset is separated into a training set and a test set. The training set is evaluated based on the cross-validation techniques. It is a resampling process that is widely used to evaluate ML models. A *k* must be specified. The *k* refers to a number of groups that must be split. The original dataset must be randomly separated into *k* group. For example, if *k* is specified as 10, ten separated group are created from the

original dataset. Cross-validation is mostly used to evaluate overall performance of ML models. Finally, a selected model is used to classify the test set. The separated unseen test set reduces the bias to the model. Figure 3.9 shows the model evaluation setting.

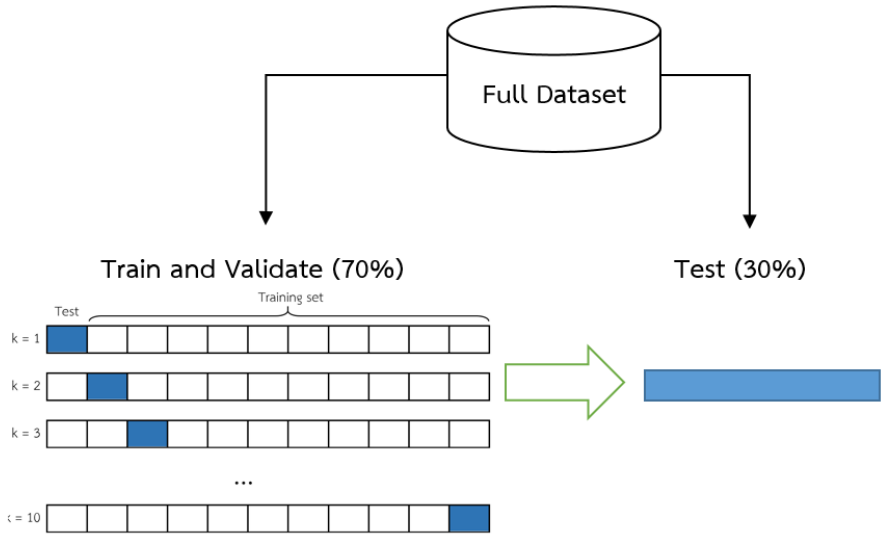


Figure 3.9 Model evaluation setting

3.6 Deployment

Both ASSC and ASDC are thoroughly studied. For the ASSC, the overall measurement and evaluation of MHLM are performed in conjunction with existing research work comparisons. The overall measurement and evaluation is also known as a subject-independent study. A subject-independent study is conducted to investigate overall performance of a particular model without testing on individual subjects. In clinical studies, a subject-dependent is necessary. The MLHM is tested in each subjects in the subject-dependent study. The subject-dependent study shows practical classification performances of the MLHM when it was used in individual subjects. For the ASDC, the detailed investigation was performed in order to identify the best ML techniques. A comparison of various ML techniques was thoroughly performed. Moreover, a feature selection process using IG was also conducted in order to identify the minimum set of best classifiers. Both proposed ASSC and ASDC are main contributions to the sleep disorder related research.

Chapter 4
Experimental Results and Discussion

4.1 Experimental Results of Proposed ASSC

In general, classification problems are studied based on the entire dataset with a predefined evaluation methods. However, clinical related classification problems are more concern in subject-based studies. Therefore, in our proposed ASSC, the model evaluations include both subject-independent and subject-dependent classifications [3]. Firstly, the evaluation of MLHM based on the subject-independent aspect was compared with ordinary DT, SVM and Sirvan Supervised Method for Sleep Staging (SSM4S) [29]. The SSM4S, a previous research work, was classifying the same dataset with selected features. Ordinary DT, SVM and the SSM4S used 30 sec epoch length. Table 4.1 shows results of the subject-independent classification on the test set.

The results are showed in $\bar{X}$±S.D. format. According to the AUC, the MHLM gained the highest value comparing DT, SVM and SS4MS in all sleep stages including the average. One of the remarks is placed to W and R. The MHLM and the SVM gained comparable AUC. For instance, in W, MHLM achieved accuracy of 0.953±0.01 (AUC = 0.920±0.26). SVM gained a slightly lower accuracy of 0.949±0.04 (AUC = 0.920±0.28). In R, MHLM achieved accuracy of 0.961±0.01 (AUC = 0.930±0.28). SVM gained similar accuracy of 0.964±0.03 (AUC = 0.930±0.22). SS4MS and DT had lower accuracy compared with MHLM and SVM in W and R. In the NREM (N1, N2 and N3), especially the N1 and N2, SS4MS reported quite low accuracy of 0.881±0.05 (AUC = 0.670±0.17), accuracy of 0.855±0.01 (AUC = 0.840±0.16), respectively. In N3, SVM gained slightly higher accuracy compared with other techniques. It had the accuracy of 0.959±0.03 (AUC = 0.935±0.22). MHLM gained slightly lower accuracy of 0.950±0.04 (AUC = 0.940±0.30). In the all of the sleep stages, MLHM gained highest average AUC. MLHM achieved accuracy of 0.942±0.02 (AUC = 0.920±0.17). The second best is SVM. It gained accuracy of 0.943±0.03 (AUC = 0.915±0.30).

The previous subject-independent classification model has been discussed in terms of the bird eye view of MHLM compared with DT, SVM and SS4MS. However, in the clinical test, the subject-dependent classification is required. The subject-dependent classification uses the discovered classification model to test on each of the subject. For this experiment, total of 100 subjects were individually tested in the subject-dependent classification. Table 4.2 shows the subject-dependent classification results. It depicts the classification performances of MLHM, DT and SVM.

In the DT ad SVM, models were evaluated in various epoch lengths including 5, 10, 15 and 30 seconds. However, the SS4MS was excluded due to the data is not available. In average of all sleep stages, MLHM gained the highest in all measures. It had the accuracy of 0.694±0.22 (AUC = 0.822±0.31). In terms of the AUC measurement, The second best is the second best classification model is DT(5ep). It achieved the accuracy of 0.688±0.21 (AUC = 0.813±0.32). The third best is SVM(5ep). It gained the accuracy of 0.631±0.22 (AUC = 0.711±0.39). Specifically, DT overcomes SVM in the same epoch length.

Table 4.1 Subject-independent classification results

Sleep Stage	Perform. Measure	MLHM	SS4MS	DT	SVM
All sleep stages	Accuracy	0.942 ± 0.02	0.902 ± 0.08	0.932 ± 0.04	0.943 ± 0.03
	Sensitivity	0.891 ± 0.27	0.752 ± 0.16	0.803 ± 0.12	0.873 ± 0.11
	Specificity	0.954 ± 0.27	0.942 ± 0.05	0.953 ± 0.03	0.962 ± 0.03
	*F*measure	0.852 ± 0.14	NA	0.784 ± 0.12	0.813 ± 0.15
	AUC	0.920 ± 0.17*	0.845 ± 0.11	0.875 ± 0.15	0.915 ± 0.30
W	Accuracy	0.953 ± 0.01	0.941 ± 0.05	0.945 ± 0.04	0.949 ± 0.04
	Sensitivity	0.891 ± 0.16	0.882 ± 0.11	0.801 ± 0.17	0.874 ± 0.17
	Specificity	0.951 ± 0.32	0.951 ± 0.05	0.953 ± 0.03	0.972 ± 0.02
	*F*measure	0.843 ± 0.07	NA	0.764 ± 0.18	0.792 ± 0.23
	AUC	0.920 ± 0.26*	0.915 ± 0.04	0.875 ± 0.20	0.920 ± 0.28*
R	Accuracy	0.961 ± 0.01	0.894 ± 0.02	0.947 ± 0.04	0.964 ± 0.03
	Sensitivity	0.892 ± 0.15	0.822 ± 0.20	0.821 ± 0.09	0.884 ± 0.07
	Specificity	0.973 ± 0.04	0.971 ± 0.04	0.964 ± 0.03	0.982 ± 0.03
	*F*measure	0.882 ± 0.06	NA	0.793 ± 0.09	0.863 ± 0.12
	AUC	0.930 ± 0.28*	0.895 ± 0.15	0.890 ± 0.21	0.930 ± 0.22*
*N*1	Accuracy	0.920 ± 0.02	0.881 ± 0.05	0.925 ± 0.03	0.938 ± 0.03
	Sensitivity	0.912 ± 0.23	0.391 ± 0.17	0.702 ± 0.13	0.882 ± 0.13
	Specificity	0.933 ± 0.08	0.952 ± 0.04	0.963 ± 0.02	0.942 ± 0.03
	*F*measure	0.721 ± 0.45	NA	0.704 ± 0.13	0.663 ± 0.21
	AUC	0.920 ± 0.17*	0.670	0.830 ± 0.22	0.910 ± 0.21
*N*2	Accuracy	0.900 ± 0.03	0.855 ± 0.01	0.903 ± 0.04	0.906 ± 0.05
	Sensitivity	0.853 ± 0.12	0.80 ± 0.15	0.82 ± 0.09	0.84 ± 0.08
	Specificity	0.952 ± 0.06	0.88 ± 0.07	0.93 ± 0.04	0.95 ± 0.03
	*F*measure	0.892 ± 0.07	NA	0.83 ± 0.10	0.85 ± 0.09
	AUC	0.900 ± 0.29*	0.840 ± 0.16	0.875 ± 0.18	0.895 ± 0.25
*N*3	Accuracy	0.950 ± 0.04	0.942 ± 0.08	0.947 ± 0.03	0.959 ± 0.03
	Sensitivity	0.921 ± 0.68	0.842 ± 0.17	0.854 ± 0.01	0.893 ± 0.09
	Specificity	0.964 ± 0.84	0.971 ± 0.04	0.973 ± 0.02	0.984 ± 0.02
	*F*measure	0.923 ± 0.06	NA	0.854 ± 0.10	0.893 ± 0.09
	AUC	0.940 ± 0.30*	0.905 ± 0.05	0.910 ± 0.33	0.935 ± 0.22

Note: *The highest AUC.

Table 4.2 Subject-dependent classification results

Sleep Stage	Perform. Measure	MLHM	DT(5ep)	DT(10ep)	DT(15ep)	DT(30ep)	SVM(5ep)	SVM(10ep)	SVM(15ep)	SVM(30ep)
All	Acc	0.694 ± 0.22	0.688 ± 0.21	0.681 ± 0.21	0.677 ± 0.20	0.678 ± 0.22	0.631 ± 0.22	0.501 ± 0.26	0.502 ± 0.26	0.501 ± 0.26
sleep	Sens	0.822 ± 0.33	0.810 ± 0.32	0.802 ± 0.32	0.776 ± 0.32	0.800 ± 0.33	0.712 ± 0.32	0.518 ± 0.40	0.520 ± 0.40	0.518 ± 0.40
stages	Spec	0.802 ± 0.33	0.782 ± 0.31	0.752 ± 0.32	0.734 ± 0.35	0.736 ± 0.33	0.682 ± 0.32	0.524 ± 0.40	0.528 ± 0.40	0.520 ± 0.40
	F	0.748 ± 0.29	0.744 ± 0.28	0.736 ± 0.27	0.734 ± 0.28	0.736 ± 0.28	0.678 ± 0.28	0.458 ± 0.35	0.458 ± 0.35	0.458 ± 0.37
	AUC	0.822 ± 0.31**	0.813 ± 0.32*	0.798 ± 0.33*	0.780 ± 0.33*	0.798 ± 0.32*	0.711 ± 0.39*	0.519 ± 0.40*	0.521 ± 0.33*	0.520 ± 0.40*
W	Acc	0.633 ± 0.24	0.633 ± 0.24	0.633 ± 0.24	0.617 ± 0.24	0.625 ± 0.24	0.444 ± 0.28	0.300 ± 0.25	0.300 ± 0.25	0.295 ± 0.25
	Sens	0.772 ± 0.33	0.771 ± 0.33	0.763 ± 0.33	0.754 ± 0.35	0.762 ± 0.35	0.472 ± 0.42	0.181 ± 0.39	0.182 ± 0.39	0.182 ± 0.39
	Spec	0.751 ± 0.33	0.757 ± 0.33	0.722 ± 0.33	0.683 ± 0.36	0.688 ± 0.35	0.462 ± 0.42	0.248 ± 0.38	0.253 ± 0.38	0.244 ± 0.39
	F	0.714 ± 0.30	0.718 ± 0.30	0.695 ± 0.29	0.698 ± 0.30	0.691 ± 0.30	0.435 ± 0.40	0.154 ± 0.33	0.154 ± 0.33	0.157 ± 0.33
	AUC	0.769 ± 0.34**	0.769 ± 0.34	0.764 ± 0.34	0.746 ± 0.35	0.757 ± 0.35	0.465 ± 0.42*	0.180 ± 0.48*	0.180 ± 0.39*	0.180 ± 0.39*
R	Acc	0.793 ± 0.22	0764 ± 0.20	0.748 ± 0.21	0.745 ± 0.22	0.740 ± 0.22	0.741 ± 0.22	0.565 ± 0.36	0.565 ± 0.36	0.566 ± 0.36
	Sens	0.904 ± 0.29	0.854 ± 0.27	0.852 ± 0.28	0.843 ± 0.29	0.833 ± 0.29	0.802 ± 0.29	0.591 ± 0.48	0.592 ± 0.48	0.591 ± 0.49
	Spec	0.891 ± 0.29	0.822 ± 0.24	0.782 ± 0.29	0.764 ± 0.29	0.753 ± 0.29	0.882 ± 0.29	0.544 ± 0.49	0.544 ± 0.48	0.552 ± 0.49
	F	0.844 ± 0.26	0.828 ± 0.24	0.811 ± 0.24	0.801 ± 0.26	0.801 ± 0.26	0.845 ± 0.26	0.553 ± 0.45	0.554 ± 0.45	0.551 ± 0.45
	AUC	0.904 ± 0.31**	0.870 ± 0.29*	0.848 ± 0.29*	0.840 ± 0.30*	0.836 ± 0.31*	0.894 ± 0.31	0.592 ± 0.39*	0.592 ± 0.49*	0.593 ± 0.49*
*N*1	Acc	0.802 ± 0.22	0.800 ± 0.22	0.801 ± 0.22	0.772 ± 0.12	0.801 ± 0.22	0.790 ± 0.20	0.732 ± 0.29	0.732 ± 0.29	0.732 ± 0.30
	Sens	0.911 ± 0.28	0.911 ± 0.28	0.912 ± 0.28	0.802 ± 0.28	0.917 ± 0.29	0.891 ± 0.26	0.822 ± 0.39	0.824 ± 0.39	0.823 ± 0.39
	Spec	0.881 ± 0.28	0.863 ± 0.28	0.872 ± 0.27	0.772 ± 0.28	0.872 ± 0.28	0.852 ± 0.27	0.844 ± 0.39	0.846 ± 0.39	0.802 ± 0.39
	F	0.852 ± 0.26	0.858 ± 0.26	0.853 ± 0.25	0.855 ± 0.26	0.853 ± 0.26	0.851 ± 0.22	0.764 ± 0.36	0.764 ± 0.36	0.761 ± 0.36
	AUC	0.911 ± 0.28**	0.908 ± 0.28*	0.910 ± 0.27	0.811 ± 0.28*	0.910 ± 0.28	0.895 ± 0.26*	0.820 ± 0.37*	0.820 ± 0.39*	0.820 ± 0.39*
*N*2	Acc	0.566 ± 0.18	0.565 ± 0.18	0.558 ± 0.18	0.546 ± 0.16	0.566 ± 0.18	0.659 ± 0.10	0.609 ± 0.16	0.613 ± 0.16	0.609 ± 0.16
	Sens	0.722 ± 0.39	0.713 ± 0.39	0.694 ± 0.39	0.682 ± 0.39	0.711 ± 0.39	0.854 ± 0.13	0.833 ± 0.37	0.844 ± 0.36	0.833 ± 0.37
	Spec	0.702 ± 0.39	0.693 ± 0.39	0.645 ± 0.39	0.673 ± 0.39	0.682 ± 0.40	0.704 ± 0.13	0.805 ± 0.36	0.814 ± 0.37	0.802 ± 0.37
	F	0.604 ± 0.32	0.605 ± 0.32	0.594 ± 0.31	0.602 ± 0.32	0.611 ± 0.31	0.782 ± 0.08	0.675 ± 0.29	0.681 ± 0.29	0.676 ± 0.30
	AUC	0.719 ± 0.39	0.711 ± 0.39*	0.688 ± 0.39*	0.719 ± 0.39	0.714 ± 0.39	0.751 ± 0.31*	0.831 ± 0.38*	0.840 ± 0.37*,**	0.832 ± 0.36*
*N*3	Acc	0.674 ± 0.22	0.676 ± 0.22	0.667 ± 0.22	0.674 ± 0.24	0.662 ± 0.22	0.519 ± 0.31	0.305 ± 0.24	0.305 ± 0.24	0.306 ± 0.24
	Sens	0.815 ± 0.34	0.814 ± 0.33	0.802 ± 0.33	0.814 ± 0.31	0.796 ± 0.33	0.553 ± 0.49	0.172 ± 0.38	0.173 ± 0.38	0.174 ± 0.38
	Spec	0.791 ± 0.34	0.792 ± 0.33	0.755 ± 0.33	0.792 ± 0.42	0.705 ± 0.34	0.523 ± 0.49	0.203 ± 0.38	0.203 ± 0.38	0.214 ± 0.38
	F	0.741 ± 0.29	0.748 ± 0.28	0.744 ± 0.28	0.734 ± 0.28	0.734 ± 0.28	0.492 ± 0.43	0.163 ± 0.34	0.153 ± 0.33	0.161 ± 0.39
	AUC	0.807 ± 0.33**	0.807 ± 0.33**	0.795 ± 0.32*	0.807 ± 0.34**	0.785 ± 0.33*	0.548 ± 0.49*	0.173 ± 0.40*	0.172 ± 0.38*	0.174 ± 0.38*

Notes: *There is statistically significant difference between the mean of AUC of the MLHM and a corresponding classification model at p-value of 0.05.
**The highest AUC.

In addition, paired-sample t-tests were conducted to compare classification models in the subject-dependent aspect. The p-value was set at 0.05. Hypotheses in

this test is to identify the difference between mean of AUC in a sleep stage of MHLM and another model. The hypotheses are set as follows.

μ_1 is the mean of AUC in a sleep stage of MHLM
μ_2 is the mean of AUC in a sleep stage of a selected model

Null hypothesis: $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis: $H_1: \mu_1 - \mu_2 \neq 0$

The full paired-sampled t-tests were performed. Tests between models that are statistically different are clearly marked in Table 4.2. Specifically, in W, MLHM is significantly different from SVM. In R, MLHM is significantly different from all other selected ML techniques except SVM(5ep). In N1, MLHM, DT(10ep) and DT(30ep) are not significantly different. SVM(15ep) apparently gained the highest AUC in N2, which is significantly different to others. In N3, MLHM, DT(5ep) and DT(30ep) has no statistically different. It is obvious that all of the classification models stated in this research work showed some degrees of classification performance reduction in the subject-dependent.

In addition to the classification evaluations, there is another vital concern in this dataset. One of the well-known imbalance measurements is Imbalance Ratio (IR). IR calculates the severity level of imbalance. IR is simply calculated from a ratio between the numbers of instances in the majority class over the minority class. Table 4.3 and Table 4.4 show the IRs of subject-independent classifications and IRs of subject-dependent classification, respectively. The depicted IRs are calculated according to sleep stages based on the Hybrid Classification Model in MLHM. Moreover, two additional indicators are added. Firstly, CHG IR and CHG AUC are the differences between average IR and average AUC of subject-dependent and subject-independent classifications. The average IRs and average AUC are calculated from 100 subjects in both subject-dependent and subject-independent classifications. For instance, the CHG IR of Layer 1 is 106.883% according to Table 4.4. The CHG IR shows the changing of IR between subject-independent and the subject-dependent. Specifically, the IR in subject-independent and subject-dependent are 3.734 and 7.725, respectively. The difference is 3.991, which is approximately doubled. CHG AUC deducts 16.413%, from 0.920 to 0.769. In Layer 2 (NREM:REM), the IR increases from 4.773 in subject-independent to 6.913 in subject-dependent. Therefore, CHG IR is increased by 44.836%. However, AUC drops from 0.930 to 0.904, which is 2.796% in Layer 2. Therefore, the changes of IR and the CHG AUC is inversely variation.

Table 4.3 Imbalanced Ratios of subject-independent Classifications

Perform. Measures	Layer 1 $W:S$	Layer 2 REM:NREM	Layer 3		
			$N1$:non-$N1$	$N2$:non-$N2$	$N3$:non-$N3$
IR	3.734	4.773	4.022	1.030	2.243
Acc (%)	0.953 ± 0.01	0.961 ± 0.01	0.920 ± 0.02	0.900 ± 0.03	0.950 ± 0.04
Sensitivity	0.891 ± 0.16	0.893 ± 0.15	0.914 ± 0.23	0.852 ± 0.12	0.921 ± 0.68
Specificity	0.953 ± 0.32	0.972 ± 0.04	0.933 ± 0.08	0.953 ± 0.06	0.962 ± 0.84
F Measure	0.843 ± 0.07	0.884 ± 0.06	0.724 ± 0.45	0.894 ± 0.07	0.921 ± 0.06
AUC	0.920 ± 0.27	0.930 ± 0.28	0.920 ± 0.27	0.900 ± 0.31	0.940 ± 0.33
CHG IR# (%)	↓ 51.663	↓ 30.956	↓ 25.587	↓ 29.112	↓ 18.908
CHG AUC# (%)	↑ 19.636	↑ 2.876	↑ 0.988	↑ 25.174	↑ 16.481

Note: #Changes of IR or AUC.

Table 4.4 Imbalanced Ratios of subject-dependent Classifications

Perform. Measures	Layer 1 $W:S$	Layer 2 REM:NREM	Layer 3		
			$N1$:non-$N1$	$N2$:non-$N2$	$N3$:non-$N3$
IR	7.725	6.913	5.405	1.453	2.766
Acc (%)	0.633 ± 0.24	0.793 ± 0.22	0.802 ± 0.22	0.566 ± 0.18	0.674 ± 0.22
Sensitivity	0.772 ± 0.33	0.903 ± 0.29	0.911 ± 0.28	0.722 ± 0.39	0.811 ± 0.34
Specificity	0.754 ± 0.33	0.892 ± 0.29	0.882 ± 0.28	0.702 ± 0.39	0.799 ± 0.34
F Measure	0.711 ± 0.30	0.843 ± 0.26	0.852 ± 0.26	0.604 ± 0.32	0.744 ± 0.29
AUC	0.769 ± 0.32	0.904 ± 0.31	0.920 ± 0.28	0.900 ± 0.34	0.940 ± 0.33
CHG IR# (%)	↑ 106.883	↑ 44.836	↑ 34.386	↑ 41.068	↑ 23.317
CHG AUC# (%)	↓ 16.413	↓ 2.796	↓ 0.978	↓ 20.111	↓ 14.149

Note: #Changes of IR or AUC.

In the subject-dependent classifications, CHG IRs are relatively high and also affects the CHG AUC, as shown in Table 4.4. A further investigation has been conducted. A data visualisation has been conducted in order to visually observe the data distribution across feature spaces. A total of randomly selected 1000 data point in O1–A2 and LOC–A2 are presented in Figure 4.1. It shows three data distribution graphs, including top, middle and bottom sections. The graphs are visualised of W(red) and S(blue), N(blue) and R(red) and N1(blue), N2(black) and N3(red) in 5, 5, and 15 s epoch lengths, respectively. In the case of high degree of overlapping of N(blue) and R(red), the SVM gained a higher classification results comparing with DT. Table 3.8 affirmed that the best selected classifier as SVM(5ep) followed by DT(5ep). In the multi-class classification of N1(blue), N2(black) and N3(red), the DT(15ep), performed on the top rank as shown in Table 3.8. The 15 s epoch length shows some degree of overlapping between N1 and N2. Therefore, in Tables 4.3 and 4.4, the classification results of N3 (AUC = 0.940±0.33) was better than the classification of N1 (AUC = 0.920±0.28) and N2 (AUC = 0.900±0.34).

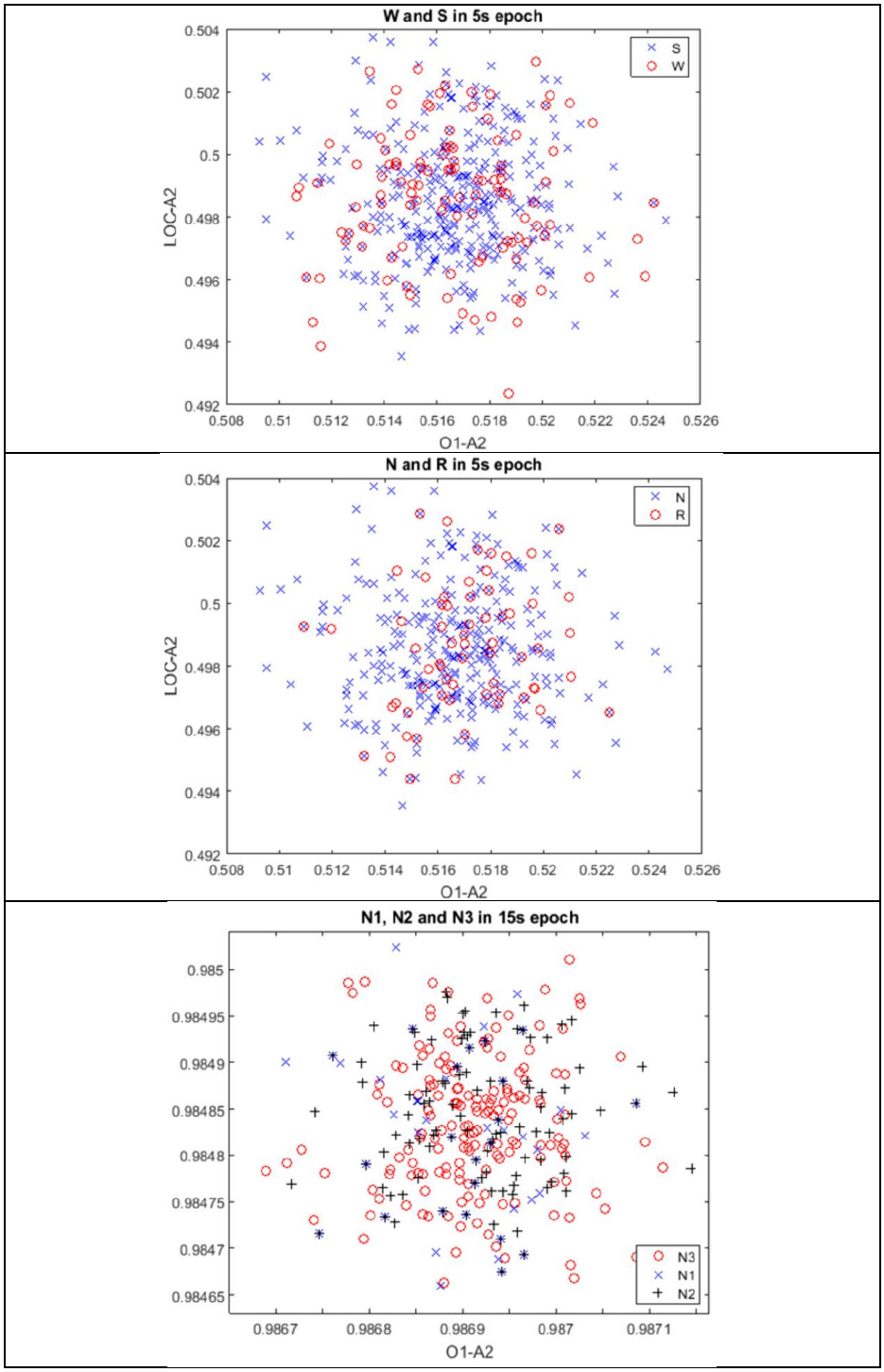


Figure 4.1 Data distribution in LOC-A2 and O1-A2

In terms of the classification performances of each classifiers in the subject-independent, a comparison between MLHM, SS4MS, DT and SVM was discussed in Table 4.1. It is obvious that the MLHM won in the overall classification performance, which included all of the sleep stages. In each of the sleep stages, the MLHM and the sole SVM gained similar classification results in W and R. However, the MLHM performed the best in the N1, N2 and N3 compared with the sole DT and SVM. In the

case of subject-specific, the classification performances of all of the classifiers dropped due to various reasons including the class imbalance and the data overlapping, which have been discussed. According to Table 4.2, the MLHM gained the highest AUC in all of the sleep stages except N2. The SVM(15ep) gained the best result. In the AASC, the overall sleep stage classification result is the main concern in order to identify the overall sleep quality.

4.2 Experimental Results of Proposed ASDC

According to Table 3.6, the average IG was calculated to identify classifying performance of each feature in sleep stages. Feature selection has been deeply investigated. Figure 3.5 depicts overall classification accuracy the selected ML techniques over the number of selected features in each sleep stage. Table 4.5 summarises overall classification performances of the selected ML techniques. The measurements include accuracy, weighted sensitivity, weighted specificity and weighted ROC. The results are calculating the overview picture of each ML techniques. The features are ranked according to the IG that were shown in Table 3.6. In Table 4.5, the calculations are performed based on the number of selected features in the ranked average IG. Specifically, the first round of calculation uses with 19 features based on the ranked average IG. The second round of calculation excluded the lowest IG feature out of the 19 features. In this case, the SNR is excluded with the average IG of 0.002. The calculation continues until the last round that includes two highest IG features, Pulse and SAO2. All of the measurements are showed in $\bar{X}$±S.D. format.

The kNN gained the best classification results in all of the sleep stages. The MLP achieved the second best position while the SVM and the kMC are at respective positions. There are two clustering based techniques used in this experiment, kNN and kMC. The kMC performs the worst in this case. The accuracies drops unacceptable in N1 and N2. In terms of selected features, the accuracies reduced dramatically at specific cut-off points. In theory, kMC classifies data points into clusters based on the nearest mean. The distance calculations between data points and clusters are calculated. The minimum distance is the indicator for putting the data points into referred clusters. Poor classification in kMC can be found in high overlapping data points, which can be in different classes. A different aspect found in the kNN. The kNN performs extremely good compared with the kMC. The kNN measures the distances between data points and their neighbours. Multiple k nearest points are used to categorise an instance into a cluster. Therefore, it is a possible reason that the kNN gains better classification results than the kMC in this experiment. The other more robust ML techniques, the SVM and the MLP. However,

both of the SVM and the MLP have steady classification performance trends. The accuracies performed by both techniques are rarely affected by the changing of number of features.

In addition, Figure 4.2 shows a data distribution in PULSE and SAO2, which are the top two highest average IGs. It is confirmed that there are degrees of data overlapping, especially in N1, N2 and N3. That is the reason that the kMC performed worst in this experiment. In R, the data overlapping is relatively low and that makes kMC classified the data more competently. Additionally, the kNN achieved a very high classification result in R. Clearly, two groups, D (blue) and P (green), are found in R. The data points are constantly distributed in R therefore the kNN is suitable for this case of experiment [44].

Table 4.5 A comparison of classification results of the kNN, kMC, SVM and MLP

Sleep stage	ML technique	Overall accuracy	Weighted sensitivity	Weighted F measure	Weighted ROC
N-1	*k*MC	42.79%±0.09	0.50±0.06	0.52±0.03	0.55±0.01
	*k*NN	96.85%±0.07	0.97±0.07	0.97±0.07	0.98±0.05
	SVM	69.13%±0.01	0.69±0.01	0.62±0.01	0.59±0.01
	MLP	74.39%±0.04	0.74±0.04	0.70±0.05	0.81±0.04
N-2	*k*MC	51.62%±0.03	0.52±0.03	0.55±0.02	0.59±0.00
	*k*NN	96.81%±0.07	0.97±0.07	0.97±0.07	0.98±0.04
	SVM	71.78%±0.01	0.72±0.01	0.68±0.01	0.65±0.01
	MLP	78.31%±0.05	0.78±0.05	0.78±0.05	0.77±0.06
N-3	*k*MC	61.64%±0.02	0.62±0.02	0.62±0.01	0.58±0.00
	*k*NN	96.24%±0.07	0.96±0.07	0.96±0.07	0.97±0.06
	SVM	72.90%±0.01	0.73±0.01	0.60±0.01	0.57±0.01
	MLP	77.09%±0.03	0.77±0.03	0.46±0.06	0.74±0.04
R	*k*MC	77.95%±0.02	0.79±0.02	0.79±0.02	0.79±0.04
	*k*NN	99.64%±0.01	1.00±0.00	1.00±0.00	1.00±0.00
	SVM	92.28%±0.02	0.92±0.02	0.92±0.02	0.94±0.01
	MLP	96.95%±0.04	0.97±0.03	0.97±0.03	0.97±0.03

According to Table 3.6, the top highest average IG was PULSE (IG = 0.335) and followed by SAO2 (IG = 0.268). The following features gained the average IGs lower than 0.13. It can be concluded that PULSE and SAO2 are the nominated featues in this experiment. In terms of the sleep disorder diagnosis and detection, PULSE and SAO2 can be measured by Photoplethysmogram (PPG). The PPG can be obtained by using a pulse oximeter. The pulse oximeter works by illuminating the skin and measuring changes in light reflection and absorption [34]. Apart from PULSE and SAO2, CANR and CHEST movements can be gathered from an ordinary ECG. EMG record the CHIN movement. Less sensor attachments to patients promote comfortability and lead to better sleep disorder and sleep classification quality.

A further deeper experiment are designed in order to promote comfortability for patients with a different number of features. Four cases of tests are designed.

Firstly, the kNN(*2f*) refers to kNN with two features, PULSE and SAO2. It requires only a pulse oximeter. Secondly, the kNN(5f) employs top five ranked features by the average IG values. Thirdly, the *opf*-kNN, the proposed model for ASDC, selects only four features including PULSE, SAO2, CANR and CHEST. These four features can be detected by two devices, the pulse oximeter and the ECG. Lastly, the *best*-kNN gains the maximum classification results with a number of selected features. The *best*-kNN selects 12, 12, 16 and 5 in N1, N2, N3, and R, respectively.

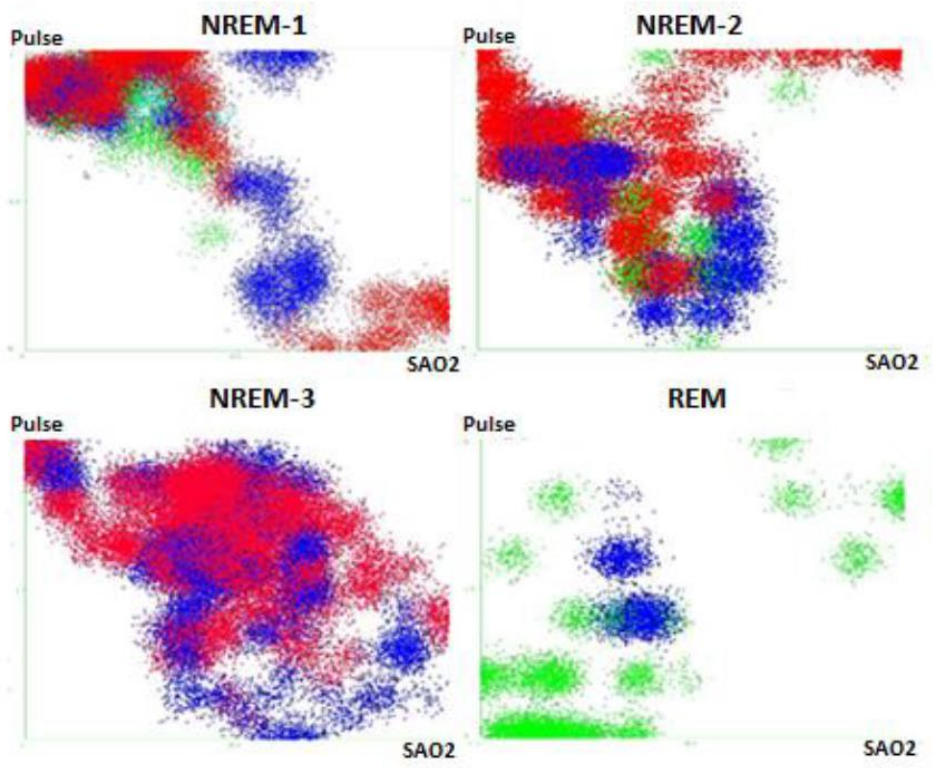


Figure 4.2 Data distribution in PULSE and SAO2

Table 4.6 shows a comparison of classification results of kNN with different selected features. The different settings include kNN(*2f*), kNN(*5f*), *opf*-kNN and *best*-kNN. The classification performance are evaluated using the 10-fold cross validation method. Certainly, the *best*-kNN is the top classifier in this experiment. It obtains the average accuracy of 99.98%+0.02. The *opf*-kNN achieves the average accuracy of 95.17%±3.91, which is the second position. The kNN(*5f*) and the kNN(*2f*) are latter positions with lesser selected features. Paired-sample t-tests were conducted to compare accuracies of classification models with the *opf*-kNN. The p-value was set at 0.05. The paired-sample t tests are conducted in order to confirm the generalization of the classification model. The result shows there is the *best*-kNN and the *opf*-kNN have no statistically significant difference ($p = 0.18$). Therefore, the overall classification accuracy of the *opf*-kNN is comparable to the *best*-kNN. The *opf*-kNN

only requires four optimal features. In addition, the *opf*-kNN performs statistically better than the kNN(*2f*).

Table 4.6 A comparison of classification results of kNN with different selected features

Sleep stage	ML technique	Overall accuracy	Weighted sensitivity	Weighted F measure	Weighted ROC
N-1	*k*NN*(2f)*	77.91%±0.03	0.78±0.03	0.32±0.01	0.77±0.01
	*k*NN*(5f)*	90.82%±0.02	0.91±0.03	0.91±0.03	0.91±0.02
	*opf-k*NN	92.67%±0.05	0.93±0.03	0.93±0.02	0.93±0.02
	*best-k*NN	99.98%±0.02	1.00±0.00	1.00±0.00	1.00±0.00
N-2	*k*NN*(2f)*	77.81%±0.04	0.78±0.03	0.23±0.04	0.78±0.04
	*k*NN*(5f)*	85.73%±0.04	0.86±0.04	0.86±0.02	0.86±0.03
	*opf-k*NN	96.64%±0.03	0.97±0.03	0.97±0.03	0.97±0.03
	*best-k*NN	99.99%±0.01	1.00±0.00	1.00±0.00	1.00±0.01
N-3	*k*NN*(2f)*	81.51%±0.04	0.82±0.03	0.36±0.04	0.80±0.04
	*k*NN*(5f)*	84.33%±0.03	0.84±0.04	0.84±0.04	0.82±0.05
	*opf-k*NN	91.40%±0.03	0.91±0.03	0.91±0.03	0.91±0.03
	*best-k*NN	99.96%±0.02	0.99±0.01	1.00±0.01	0.92±0.02
R	*k*NN*(2f)*	97.75%±0.03	0.98±0.03	0.98±0.02	0.98±0.03
	*k*NN*(5f)*	99.98%±0.02	1.000±0.01	1.000±0.01	1.000±0.01
	*opf-k*NN	99.98%±0.02	1.000±0.01	1.000±0.01	1.000±0.01
	*best-k*NN	100.00%±0.00	1.000±0.00	1.000±0.00	1.000±0.00
AVERAGE	*k*NN*(2f)*	83.75%±9.49	0.84±0.10	0.23±0.16	0.83±0.10
	*k*NN*(5f)*	90.22%±7.08	0.90±0.07	0.90±0.07	0.90±0.08
	*opf-k*NN	95.17%±3.91	0.95±0.04	0.95±0.04	0.95±0.04
	*best-k*NN	99.98%+0.02	0.99±0.01	1.00±0.00	0.98±0.04

Chapter 5
Conclusions

5.1 Conclusions

Both ASSC and ASDC become main areas in sleep disorder research works. Efficient ASSC and ASDC are automated or semi-automated methods that assist sleep specialists in interpreting bio-signals from the PSG. The correct interpretations lead to properly sleep stage and sleep disorder classification. Both ASSC and ASDC are frequently constructed from mathematical, statistical and ML models. Our proposed ASSC model introduced Multi-Layer Hybrid ML Model (MLHM) that comprised of a hybrid classification model. A newly developed multi-layer architecture was also embedded. The hybrid classification model consisted of two baseline ML techniques, Decision Tree (DT) and Support Vector Machine (SVM). In addition, different epoch lengths were deeply studied and selected in order to gain better classification results. Experiments were set up to perform in both subject-dependent and subject-independent. The MLHM achieved acceptable classification results, 94.20%±0.02.

The objective of our proposed ASDC model was to classify sleep-related syndromes and sleep disorders with optimal features. The optimal features were ranked using the average Information Gain (IG). The selected optimal features were PULSE, SAO2, CANR and CHEST. The investigation of different ML techniques was performed in order to select the best classifier. In general, the highest classification accuracy for our experiment was the kNN. The latter techniques were MLP, SVM and kMC, respectively. An additional experiment was conducted to compare the kNN with various combinations of number of features. A selected optimal feature kNN, opf-kNN, used gained the accuracy of 95.17%±3.91. The opf-kNN has no statistically different comparing to the best-kNN, the best kNN classifier with selected number of features. This study not only attempting to locate the best classifier but also concerning about the use of high-priced medical equipment and the patient comfortability.

5.2 Discussions and Future Works

Apart from the classification results, there are a number of interesting points to be discussed. Firstly, the feature extraction and selection process is still developing. Various types of feature extraction and selection techniques are being proposed. There is no final decision towards the best feature extraction and selection technique. One of the reasons is collected bio-signals from different PSG

can be very different in terms of features and scales. Therefore, the feature extraction and selection techniques can be very varied. Most of the research works are targeting to achieve highest classification results. However, in our research works, selections of optimal features attempt to balance between the best classification results and patient comfortability.

Secondly, continuing from the feature extraction and selection, selected epoch length is another factor to concern. The standard epoch length in the PSG is 30 seconds interval. The reduction or expansion of the epoch length can be surplus towards the classification results. Different epoch lengths that has been used in many research works including 5s, 10s, 15s and 30s. One of our research works has identified different epoch lengths that are suitable for different sleep stages.

Thirdly, the model evaluation is another concerning points. The subject-independent context represents overview classification performance of a particular classification model. In general, a PSG dataset is divided into a training set and a test set. The test set usually is set to be unseen to a classification model until the final evaluation is performed. However, in the medical field, the subject-dependent is also interested by practitioners. The subject-dependent context represents how a classification model performs on each subject.

In conclusion, there are various points to be grown in both ASSC and ASDC frameworks including selected feature extraction and selection process, ML techniques used, classification model evaluations especially the subject-dependent context.

References

[1] World Health Organization (WHO), “Cardiovascular Disease, Fact sheet N°317 [Online],” 2015.

[2] World Health Organization (WHO), “Non Communicable Diseases,” 2018. [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases. [Accessed 1 May 2018].

[3] Assessment IoMRaT, “Literature Review: The current situation and care model of non-communicable diseases,” Nonthaburi: Ministry of Public Health, Institute of Medical Research and Technology Assessment, 2014.

[4] O. Kocak, T. Bayrak, A. Erdamar and L. Ozparlak, “Automated Detection and Classification of Sleep Apnea Types Using Electrocardiogram (ECG) and Electroencephalogram (EEG) Features,” *Advances in Electrocardiograms - Clinical Applications,* pp. 211-230, 2012.

[5] W. Szelenberger and C. Soldatos, “Sleep disorders in psychiatric practice,” *World Psychiatry,* vol. 4, no. 3, pp. 186-190, 2005.

[6] B. Jafari and V. Mohsenin, “Polysomnography,” *Clinics in Chest Medicine,* vol. 31, no. 2, pp. 287-297, 2010.

[7] J. Haba-Rubio and J. Krieger, “Evaluation instruments for sleep disorders: a brief history of polysomnography and sleep medicine,” in *Introduction to modern sleep technology*, Netherland, Springer, 2012, pp. 19-31.

[8] R. Berry, Fundamentals of sleep medicine, Saunders: Elsevier, 2012.

[9] O. Kocak, T. Bayrak, A. Erdamar, L. Ozparlak, Z. Telatar and O. Erogul, “Automated Detection and Classification of Sleep Apnea Types Using Electrocardiogram (ECG) and Electroencephalogram (EEG) Features,” *Advances in Electrocardiograms - Clinical Applications,* pp. 211-230, 2012.

[10] W. Szelenberger and C. Soldatos, “Sleep disorders in psychiatric practice,” *World Psychiatry,* vol. 4, no. 3, pp. 186-190, 2005.

[11] American Academy of Sleep Medicine (AASM), “Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force,” *Sleep,* vol. 22, no. 5, pp. 667-689, 1999.

[12] A. Z. Azarbarzin, “Snoring sounds' statistical characteristics depend on anthropometric parameters,” *Journal of Biomedical Science and Engineering,*

vol. 5, no. 245-254, 2012.

[13] D. Moser, P. Anderer, G. Gruber, S. Parapatics, E. Loretz, M. Boeck, G. Kloesch, E. Heller, A. Schmidt, H. Danker-Hopfe, B. Saletu, J. Zeitlhofer and G. Dorffner, "Sleep Classification According to AASM and Rechtschaffen & Kales: Effects on Sleep Scoring Parameters," *Sleep,* vol. 32, no. 2, pp. 139-149, 2009.

[14] H. Danker-Hopfe, D. Kunz and G. Gruber, "Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders," *Journal of Sleep Research,* vol. 13, pp. 63-69, 2004.

[15] M. YAN, X. XU, . Z. HUANG, Y. Ming-Hui, . Y. URADE and . W. QU, "Selection of optimal epoch duration in assessment of rodent sleep–wake profiles," *Sleep and Biological Rhythms,* vol. 9, no. 1, pp. 46-55, 2011.

[16] C. Shearer, "The CRISP-DM model: the new blueprint for data mining," *Journal of Data Warehousing,* vol. 5, pp. 13-22, 2000.

[17] D. T. Larose, Discovering knowledge in data : an introduction to data mining, Hoboken, New Jersey: John Wiley & Sons, Inc., 2005.

[18] T. Wongsirichot and A. Hanskunatai, "A Comparative Investigation of PSG Signal Patterns to Classify Sleep Disorders Using Machine Learning Techniques," in *ICIC 2015. Lecture Notes in Computer Science*, 2015.

[19] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Technique, Waltham, MA: Morgan Kaufmann Publishers, 2012.

[20] B. Tay, J. K. Hyun and S. Oh, "A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained from Diffusion Tensor Images," *Computational and Mathematical Methods in Medicine,* pp. 1-8, 2014.

[21] C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* vol. 2, pp. 121-167, 1998.

[22] A. Šimundić, "Measures of diagnostic accuracy: basic definitions," *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine,* vol. 19, no. 4, pp. 203-211, 2009.

[23] N. Chawla, N. Japkowicz and A. Kotcz, "Imbalanced Data Sets," in *International Conference of ICML Workshop Learn*, 2003.

[24] Y. Ye, K. Yang, J. Jiang and B. Ge, "Automatic sleep and wake classifier with heart rate and pulse oximetry: Derived dynamic time warping features and logistic model," in *Annual IEEE in Systems Conference (SysCon)*, Orlando, FL, USA, 2016.

[25] S. Khalighi, T. Sousa, J. M. Santos and U. Nunes, "ISRUC-Sleep: a comprehensive public dataset for sleep researchers," *Computer Methods and Programs in*

Biomedicine, vol. 124, pp. 180-192, 2015.

[26] A. Bharadwaj and S. Minz, "Hybrid approach for classification using support vector machine and decision tree," in *International Conference of Advances in Electronics, Electrical and Computer Science Engineering (EEC 2012)*, 2012.

[27] T. Lajnefa, S. Chaibia, P. Rubyb, P. Aguerab , J. Eichenlaubc, M. Sameta, A. Kachouria and K. Jerbi, "Learning machines and sleeping brains: Automatic sleep stage classification using decision tree multi-class support vector machines," *Journal of Neuroscience Methods,* vol. 250, pp. 94-105, 2015.

[28] K. Aboalayon and M. Faezipour, "Multi-class SVM based on sleep stage identification using EEG signal," in *Health Innovations and Point-of-Care Technologies Conference*, 2016.

[29] S. Khalighi, T. Sousa, G. Pires and U. Nunes, "Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels," *Expert Systems with Applications,* vol. 40, no. 17, pp. 7046-7059, 2013.

[30] T. Cargone, L. Marrero, J. Weiss, M. Hiatt and T. Hegyi, "Heart Rate and Oxygen Saturation Correlates of Infant Apnea," *Journal of Perinatology,* vol. 19, no. 1, pp. 44-47, 1999.

[31] J. Sun, X. Li, J. Guo, F. Han and H. Zhang, "Identification of Obstructive Sleep Apnea Syndrome by Ambulatory Electrocardiography: Clinical Evaluation of Time-domain and Frequency-domain Analyses of Heart Rate Variability in Chinese Patients," *Cell Biochem Biophys,* vol. 59, pp. 165-170, 2011.

[32] F. Chapotot and G. Becq, "Automated sleep–wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules," *International Journal of Adaptive Control and Singal Processing,* vol. 24, no. 5, 2009.

[33] L. Almazaydeh, M. Faezipour and K. Elleithy, "A Neural Network System for Detection of Obstructive Sleep Apnea Through SpO2 Signal Features.," *International Journal of Advanced Computer Science and Applications,* vol. 3, no. 5, pp. 7-11, 2012.

[34] J. Behar, A. Roebuck, M. Shahid, J. Daly, A. Hallack and N. Palmius, "SleepAp: An Automated Obstructive Sleep Apnoea Screening Application for Smartphones," *Computing in Cardiology,* vol. 2013, pp. 257-260, 2013.

[35] T. Wongsirichot, N. Iad-ua and J. Wibulkit, "A Snoring Sound Analysis Application using K-Mean Clustering Method on Mobile Devices," in *the 9th International Conference on Computer Recognition Systems CORES 2015*, Wroclaw, Poland,

2016.

[36] S. Khalighi, T. Sousa, J. Santos and U. Nunes, “ISRUC-Sleep: A comprehensive public dataset for sleep researchers,” *Computer Methods and Programs in Biomedicine,* vol. 124, pp. 180-192, 2016.

[37] N. M. Incorporated, “Natus Medical Incorporated,” 2012. [Online]. Available: http://www.natus.com/. [Accessed 12 June 2015].

[38] M. Albayrak and E. Koklukaya, “The detection of an epileptiform activity on EEG signals by using data mining process,” *E-Journal of New World Sciences Academy,* vol. 4, pp. 227-250, 2009.

[39] T. Sousa, D. Oliveria, S. Khalighi, G. Pires and Nunes U, “Neurophysiologic and statistical analysis of failures in automatic sleep stage classification,” 2012.

[40] T. Wongsirichot and A. Hanskunatai, “A comparative investigation of PSG signal patterns to classify sleep disorders using machine learning techniques,” in *Lecture Notes in Computer Sciences*, vol. 9225, 2015, pp. 510-521.

[41] T. Sousa, A. Cruz, S. Khalighi, G. Pires and U. Nunes, “A two-step automatic sleep stage classification methods with dubious range detection,” *Computers in Biology in Medicine,* vol. 59, pp. 42-53, 2015.

[42] S. Khalighi, T. Sousa, D. Oliverira, G. Pires and U. Nunes, “Efficient feature selection for sleep staging based on maximal overlap discrete wavelet transform and SVM,” in *Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)*, 2011.

[43] Y. Yang and J. Pedersen, “A comparative study on feature selection in text categorization,” in *ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning*, San Francisco, CA, USA, 1997.

[44] C. Domeniconi, J. Peng and D. Gunopulos, “Locally adaptive metric nearest-neighbor classification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 9, 2002.

[45] M. Ramaswami and R. Bhaskaran, “A study of feature selection techniques in educational data mining,” *Journal of Computer,* vol. 1, pp. 7-11, 2009.

[46] J. García and R. Mollineda, “On the effectiveness of preprocessing methods when dealing with different levels of class imbalance,” *Knowledge-based Systems,* vol. 25, pp. 13-21, 2012.

[47] T. Wongsirichot and A. Hanskunatai, “A Classification of Sleep Disorders with Optimal Features using Machine Learning Techniques,” *Journal of Health Research,* vol. 31, no. 3, pp. 209-217, 2017.

Author Biography

Name	Mr. Thakerng Wongsirichot
Date of Birth	22 February 1981
Address	311/21 Moo 6, Thung Yai, Hat Yai, Songkhla, 90110
Education	2003 Master of Information Systems (with Distinction), University of Wollongong, Australia
	2001 Bachelor of Commerce (Electronic Commerce and Business Information System), University of Wollongong, Australia

Academic Publications

1. T. Wongsirichot and A. Hanskunatai, “A comparative investigation of PSG signal patterns to classify sleep disorders using machine learning techniques,” in Lecture Notes in Computer Sciences, vol. 9225, 2015, pp. 510-521.
2. T. Wongsirichot and A. Hanskunatai, “A Classification of Sleep Disorders with Optimal Features using Machine Learning Techniques,” *Journal of Health Research,* vol. 31, no. 3, pp. 209-217, 2017.
3. T. Wongsirichot and A. Hanskunatai, “A Multi-Layer Hybrid Machine Learning Model for Automatic Sleep Stage Classification,” *Biomedical Engineering: Applications, Basis and Communications,* vol. 31, no. 6, pp. 1-13, 2018.

Appendix

Appendix A

A Comparative Investigation of PSG Signal Patterns to Classify Sleep Disorders Using Machine Learning Techniques

Thakerng Wongsirichot(✉) and Anantaporn Hanskunatai

Department of Computer Science Faculty of Science,
King Mongkut's Institute of Technology Ladkrabang (KMITL),
Bangkok, 10520, Thailand
thakerng.w@gmail.com, ksananta@kmitl.ac.th

Abstract. Patients with Non-Communicable Diseases (NCDs) are increasing around the globe. Possible causes of the NCDs are continuously being investigated. One of them is a sleep disorder. In order to detect specific sleep disorders, the Polysomnography (PSG), is necessary. However, due to the lack of the PSG in many hospitals, researchers attempt to discover alternative approaches. This article demonstrates comparisons of sleep disorder classifications using machine learning techniques. Three main machine learning techniques have been compared including Classification And Regression Tree (CART), *k*-Mean Clustering (KMC) and Support Vector Machine (SVM). The SVM achieves the best classification results in NREM-1 and NREM-2. The CART performs superior in NREM-3 and REM. Implications in terms of medical diagnosis, there are two main selected features, SaO2 and Pulse, based on the CART in all of the sleep stages. The features may be pieces of evidences to predict various types of sleep disorders.

Keywords: Sleep disorders · Classification · CART · *K*-Mean clustering · SVM

1 Introduction

One third of our lifetimes are allocated for one of the most vital activities, sleeping. Wakefulness has been subsided during sleep intervals with minimum apparent physical activities. The Central Nervous System (CNS) plays a significant role in controlling all human's activities even in the sleeping periods. Quality of sleep is considered as a vital property of human beings unless some symptoms may be recognized, especially the Non-Communicable Disease (NCD). Statistically, almost 1 of 2 adults (more than 133 million) in the United States live with at least one NCD. There are a number of clinical studies show that adults with NCDs such as hypertension and cardiovascular diseases are dramatically increasing worldwide [1]. The NCDs are caused by various factors such as lacks of exercises, genetic inheritances and mutations, prolonged use of some medications, etc. NCD patients' conditions are steadily degrading if medications or treatments are not properly provided. In order to discover appropriate medications or treatments, deepen diagnostic techniques are

D.-S. Huang et al. (Eds.): ICIC 2015, Part I, LNCS 9225, pp. 510–521, 2015.
DOI: 10.1007/978-3-319-22180-9_50

possibly appointed. With advancements of medical diagnostic technologies, a number of NCDs and symptoms are deeply studied in order to accomplish their causes and discover suitable treatments.

According to the American Academy of Sleep Medicine (AASM) standards, a sleep cycle contains four stages that are separated into two phrases namely, the Rapid Eye Movement (REM) and the Non-Rapid Eye Movement (NREM). There is only one stage in REM and three stages in NREM including stage 1, 2 and 3 [2]. A sleep cycle initiates when eye lids are closed (NREM-1), follows by a light sleep episode (NREM-2), and a deep sleep episode (NREM-3). The REM phrase is entered immediately after the NREM-3. An identification of sleep stages is a key main factor to detect sleep disorders [3].

Clinically, there are eight types of sleep disorders and may be classified into two main categories so called Dyssomnias and Parasomnias [4]. The sleep disorders may develop to serious sleep disorders that is the sleep apnea. One of the most encountered sleep apnea is Obstructive Sleep Apnea (OSA). The OSA is caused by the obstructions of the upper airway and series of repetitive pauses of breathing (apnea) during sleep. Possible causes of the OSA include overweight, short jaw structure, etc. Another type of sleep disorder, the Central Sleep Apnea (CSA), is similar to the OSA. However, the CSA is caused by abnormal brain activities [5]. The OSA and CSA are obviously found in severe patient cases. Most of the researchers attempt to identify only the OSA-related problems. However, there are other related sleep disorders that also require in-depth attentions and researches, especially the potential OSA/CSA patients.

2 Related Works

Researchers in various research fields including medicines, biomedicines and computer sciences attempt to investigate information beneath gigantic amount of recorded data from the PSG test. A group of researchers performed analyses using statistical analytical techniques such as regression analyses, autoregressive techniques, etc. On the other hand, others utilise machine mining techniques to conduct the analysis such as the Artificial Neuron Network (ANN), the Support Vector Machine (SVM), etc. Due to the vast amount of raw data have been collected in each PSG test, practically only subsets of potential variables, which are in EEG, EKG, ECG or a combination of these signals, are selected for analyses. Additionally, the PSG test is one of the dedicated systems that has to be performed in a formal sleep laboratory. Due to high setup and maintenance costs, the PSG test is rarely available in many hospitals.

Over the past few decades, there are a number of targeted investigations of the OSA and the CSA. A research work performed a pattern analysis in order to discriminate the CSA and the OSA using only the EEG recorded data. The researchers selected series of signals from C3-A2 and C4-A1 nodes wherein the sleep stage 2 per se. The Feed-Forward Neural Network (FFNN) was used to classify three studied groups, which have been preclassified by sleep specialists. Specifically, there are ten subjects with the CSA, ten subjects with the OSA, and ten healthy subjects. The EEG

synchronisation methods, both the Coherence Function (CF) and the Mutual Information (MI), have been formulated to discriminate the CSA and the OSA cases. The classification result is 93.3 % accuracy [6].

Rather than using the partial EEG signal to classify the OSA and the healthy subjects, a group of researchers gathered and analysed the ECG signals. Clinically, there are common irregular ECG patterns, specifically based on the Heart Rate Variability (HRV), that are associated with apnea. A presence of bradycardia, the heart rate is below 60 bpm, follows by tachycardia, the heart rate is over 100 bpm. In order to conduct the HRV analysis, which is non-stationary, the wavelet decomposition has been engaged. In addition to the analysis of the HRV, the QRS complex also represents a correlation between the apnea and its pattern. The ECG-derived respiration (EDR) is retrieved in order to investigate the attenuation of respiratory effort. Therefore, the combination of HRV and EDR have been selected as classifying parameters using the SVM technique. 83 subjects have been used to develop the classification algorithm in this study together with 42 test cases from three sources. The collected data has been initially processed using the Wavelet Decomposition technique. It generated 14 levels with the Daubechies wavelets. The SVM has been performed in order to conduct the binary classification. The final result showed a promising result of 92.85 % accuracy on the independent tests with Cohen's K value of 0.85 [7].

In general, the OSA subjects usually have daytime sleepiness occurrences even they had full night sleeps. The researchers observe the OSA subjects that are related to daytime sleepiness. A research work collected data from three different groups. There are five untreated subjects, four narcoleptic subjects, and six healthy subjects. The interesting point of this research work is pupillometry data have been considered as a classifier along with the EEG signal. Specifically, the EEG signals have been selected only the signals from C3-A2, O1-A2 and P3-O1, which relate to eye movements. The NN technique, specifically the ART2 NN algorithm, has been selected as a main classifier between the OSA and the healthy subjects. The result shows 91 % classification accuracy [8].

Since 2002, there are a number of researchers interested in studies of snoring sound may relate to the NCDs. A report showed approximately 20–40 % of adults snore whilst asleep [9] However, not all of the snorers have or will eventually encounter the NCDs. Out of the medical observations and diagnostics, our previous work attempted to find other possible variables that are able to predict sleep disorder episodes. We also attempted to analyse snoring sound patterns with mobile devices using the modified *k*-Mean Clustering technique. With our experiment test, 74.70 % instances has been correctly classified. It implies that only the snoring sound patterns may not be sufficient for the classifications [10]. According to the reviewed related works, the sample sizes of the studied subjects are limited due to the availabilities of the data, which are collected from actual clinical sleep test laboratories. All of the studied subjects are extracted from the full PSG recordings. In addition to the sample sizes, most of the researches targeted to only severe cases of the OSA or the CSA. However, people with apnea risks are not thoroughly studied.

3 Methods

Our study investigates not only severe sleep apnea cases per se. It also includes mild or non-sleep apnea cases with anonymity. A set of full PSG recordings from five OSA cases were digitally extracted from a PSG recording software application, which installed at the Songklanagarin Hospital, Thailand. The study has been approved by the hospital's director with explicit attentions to the patient confidentiality. The main objective is to apply the DT, specifically the Classification and Regression Trees (CART), and the *k*-Mean Clustering (KMC) techniques to classify frequently found sleep disorder patterns. A set of 19 original explanatory variables has been stamped, which are predefined by the PSG machine [10]. (Table 1).

The variables are categorised into EEG, ECG, EKG, and other movement detections. The recorded data are separated according to sleep stages and excluded the normal sleep patterns. Four sleep disorders are manually identified by sleep technicians including Oxygen Desaturation (D), Hypopnea (H), PLMD (P), and IPLMD (I) episodes. Specific skills are required for the identification processes. Due to the gigantic amount of data, a simple technique that is able to represent possible decision paths and chance events. Moreover, the CART technique is able to act as a preliminary feature selector choosing possible nominated variables. On the other hand, the KMC technique is selected, which has been used in our previous research work, for a comparison purpose [11]. Additionally, the SVM, a well-known classification technique [7] has been tested on this clinical dataset.

Algorithm 1 CART (rpart function in R) [12]
Require: *D* is a set of PSG records.
attr_list is a set of selected attributes.
rpart() is the attribute selection method in the R (rattle package).

```
create a node N //as an initial node
if tuples in D are all the same as C, then
  return N as a leaf node labelled with the class C;
if attr_list is empty then
  return N as a leaf node labelled with the majority
  class in D;
apply rpart(D, attr_list) to find the "best"
splitting_criterion;
label Node N with splitting_criterion;
if splitting_atrribute is discreate-valued and
  Multiway splits allowed then
    attr_list ← attr_list - splitting_attribute
for each outcome j of splitting_criterion
  let Dj be the set of data records in D satisfying
    outcome j
  if Dj is empty then
    attach a leaf labelled with D to node N
  else attach the node returned by
  Generate_CART_decision_tree(Dj, attr_list) to node N
endfor
return N
```

Table 1. Variables from the PSG Signals

Variable	Description	Variable	Description
C3-A2	Monopolar EEG at C3-A2	FLOW	Mouth Airflow
C4-A1	Monopolar EEG at C4-A1	CHEST	Chest movement
F3-A2	Bipolar EEG at F3-A2	ABDOMEN	Abdomen movement
F4-A1	Bipolar EEG at F4-A1	LAT	Left Anterior Tibialis
O1-A2	Monopolar EEG at O1-A2	RAT	Right Anterior Tibialis
O2-A1	Monopolar EEG at O2-A1	EKG	Electrocardiography
CHIN	Chin Movement	Pulse	Pulse
LOC	Left Outer Canthus	SaO2	Saturation level of oxygen in haemoglobin
ROC	Right Outer Canthus	Snore	Amplitude of snoring sounds
canular	Nasal Cannula		

The CART technique is employed to classify the dataset into meaningful tree-like structures. There are three main parameters including *D*, *attr_list*, and *rpart()*. Specifically, the D is a data partition of the extracted PSG dataset. The *attr_list* is a set of selected attributes of the dataset. The *rpart()* is an attribute selection method in R. It selects a set of best discriminated attributes in order to classify the dataset into corresponding classes [12]. With the same dataset, the KMC has been performed to partition observations into potential sleep disorder classes (k). In general, the initial dataset has been loaded for a training purpose. According to Algorithm 2, a number of sleep disorder classes in each of the sleep stages are predefined by sleep technicians. The classes are the number of clusters k in the KMC. An initial set of centroids is calculated with continuously adjustments throughout the process. The final set of centroids are referenced in order to evaluate the testing dataset. The Euclidean distance *(argminDistance(D_i, C_k))* is used to measure the differences between an element with the closest centroid. Final decisions of classifications are represented with the minimum distance between the element and its corresponding centroid [12].

Algorithm 2 *k*-Mean Clustering [12]
Require: *D is a set of PSG records.*
k is a number of clusters.

1: **set** k based on predefined sleep disorder classes
2: **let** C_k is a centroid of each clusters
3: **repeat**
4: **assign** D_i to a cluster to which the record is the most similar using argminDistance(D_i, C_k)
5: **update** C_k
6: *MaxIter*++
7: **until** *MaxIter* is 500 // a number of iterations

4 Results

All of the computational analyses have been successfully performed by the Rattle package in R. The dataset has been divided into two portions, training and testing datasets, for our analyses with the ratio of 70:30, respectively. Table 2 represents the CART analysis results.

Four main performance evaluation of the measures have been selected to evaluate each of the techniques according to the sleep stages. The measures include Accuracy, Precision, Sensitivity, Specificity, and F-Measure. The measures are calculated from incremental counts of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) in confusion matrices. The followings are the formulae of the measures [13].

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \tag{1}$$

$$Precision = \frac{|TP|}{|TP| + |FP|} \tag{2}$$

$$Sensitivity = \frac{|TP|}{|TP| + |FN|} \tag{3}$$

$$Specificity = \frac{|TN|}{|TN| + |FP|} \tag{4}$$

$$F - Measure = \frac{(2 \times |TP|)}{(2 \times |TP|) + |FP| + |FN|} \tag{5}$$

Specifically, confusion matrices are constructed according to sleep stages on both of the analysis methods, the CART, the KMC and the SVM. D, H, I, and P are the classes of the Oxygen Desaturation, Hypopnea, IPLMD, and PLMD, respectively. Furthermore, $\hat{D}, \hat{H}, \hat{I}$ and $\hat{P}$ are the representations of classified classes corresponding to D, H, I, and P, respectively. Due to the confusion matrices are based on multi-classes classification, each of the measures is compared a class with the rest of the classes. Tables 3 and 4 represent the confusion matrices, the sleep disorder episode classification results, of NREM-1, NREM-2, NREM-3 and REM using the CART technique. In order to clarify the evaluation measures based on the confusion matrices, Table 4 in

Table 2. Features Selected by the CART technique

Sleep Stage	Observation (n)	Tree Construction Variables
NREM-1	86072	Canular, Pulse, SaO2
NREM-2	83274	Abdomen, Chest, Flow, Pulse, SaO2
NREM-3	85164	Abdomen, LOC, Pulse, ROC, SaO2
REM	14958	O2.A1, Pulse, SaO2

Table 3. Confusion Matrices of the NREM-1 and NREM-2 using the CART Technique

NREM-1	$\hat{D}$	$\hat{H}$	$\hat{I}$	$\hat{P}$	NREM-2	$\hat{D}$	$\hat{H}$	$\hat{P}$
D	5457	11398	5399	5331	D	19903	7442	0
H	3359	24801	11474	15522	H	5603	47402	0
I	0	80	0	258	P	330	2594	0
P	129	429	137	2298				

Table 4. Confusion Matrices of the NREM-3 and REM using the CART Technique

NREM-3	$\hat{D}$	$\hat{H}$	REM	$\hat{D}$	$\hat{P}$
D	37473	84	D	5537	0
H	77	84030	P	9	9389

the REM stage, the TP value is the corrected classified D class as the $\hat{D}$, class, which is 5537. The FP value is the P class that has been classified as $\hat{P}$, which is 9 (Bottom left in the matrix). The FN value is the D class that has been classified as, which is 0 (Top right in the matrix). The TN value is the P class that has been classified as $\hat{P}$, which is 9389. Additionally, the calculations of TP, TN, FP and FN are performed separately according to each of the base class. For example, if class D is considered, the TP, TN, FP and FN are based on only class D.

Tables 5 and 6 show the performance evaluation results in each of the sleep stages. In NREM-1, there are four determined classes, D, H, I and P. The classification of the $H - \hat{H}$ achieves highest F-Measure value of 0.54. In NREM-2, there are three determined classes, D, H and P. The classification of the $H - \hat{H}$ achieves highest F-Measure value of 0.86. However, in the NREM-3 and REM, all of the F-Measure values reach the maximum of 1.00, which mean the majority of the instances are correctly classified.

Figure 1 represents a tree structure of the REM stage using the CART technique. The CART technique implements a binary classification [11] There are three selected variables including O2.A1, Pulse and SaO2, which have been mentioned in Table 2. In this sleep stage, there are two classes, D and P. At the beginning of the tree, the condition of SaO2 > = 94 is set based on the root node calculation. If the SaO2 > = 94

Table 5. Performance Evaluation Results of NREM-1 and NREM-2 using the CART Technique

Performance Measure	NREM-1				NREM-2		
	D-$\hat{D}$	H-$\hat{H}$	I-$\hat{I}$	P-$\hat{P}$	D-$\hat{D}$	H-$\hat{H}$	P-$\hat{P}$
Accuracy	0.702	0.509	0.798	0.747	0.839	0.812	0.965
Precision	0.610	0.676	0.000	0.098	0.770	0.825	–
Sensitivity	0.198	0.450	0.000	0.768	0.728	0.894	0.000
Specificity	0.941	0.615	0.801	0.746	0.894	0.668	1.000
F-Measure	0.30	0.54	0.00	0.00	0.75	0.86	0.00

Table 6. Performance Evaluation Results of NREM-3 and REM using the CART Technique

Performance Measure	NREM-3		REM	
	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	P-$\widehat{P}$
Accuracy	0.999	0.999	0.999	0.999
Precision	0.998	0.999	0.998	1.000
Sensitivity	0.998	0.990	1.000	0.999
Specificity	0.999	0.998	0.999	1.000
F-Measure	1.00	1.00	1.00	1.00

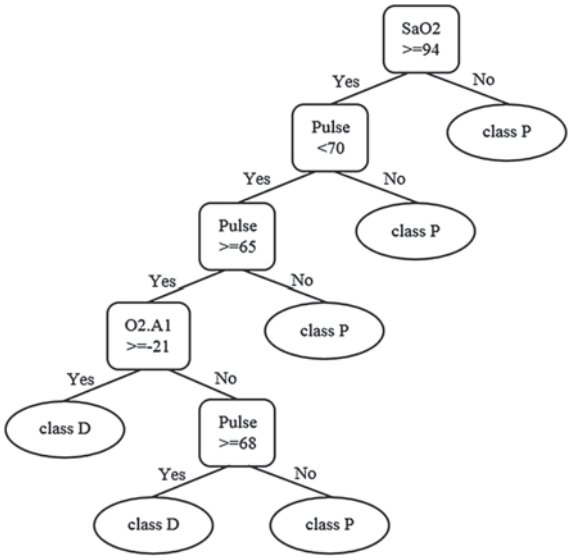


Fig. 1. A Tree Structure of the REM Stage using the CART Technique

is true, the latter condition is Pulse < 70. If the SaO2 >= 94 is false, the element is classified to the class P. The process is continuously performed until all of the instances are classified into classes.

On the other hand, the KMC has been performed with the k clusters based on actual known cluster sizes in each of the sleep stages. For example, the NREM-1 has four types of presented sleep disorders so that the k is set to four.

Tables 7 and 8 represent the confusion matrices, classifications of the sleep order episode results, of the NREM-1, NREM-2, NREM-3 and REM using the KMC technique. Overall performance evaluation results show that the KMC technique also performs relatively acceptable in the REM stage, in Tables 9 and 10. Specifically, the $P - \hat{P}$ and $H - \hat{H}$ are best in REM and NREM-1 in some classified classes, respectively. On the other hand, the $I - \hat{I}$ in NREM-1, is totally incorrectly classified.

Table 7. Confusion Matrices of NREM-1 and NREM-2 using the KMC Technique

NREM- 1	$\widehat{D}$	$\widehat{H}$	$\widehat{I}$	$\widehat{P}$	NREM-2	$\widehat{D}$	$\widehat{H}$	$\widehat{P}$
D	8272	19049	0	264	D	13037	14308	0
H	409	54747	0	0	H	13321	36479	3205
I	0	0	0	338	P	2441	177	306
P	129	629	0	2235				

Table 8. Confusion Matrices of the NREM-3 and REM using the KMC Technique

NREM-3	$\widehat{D}$	$\widehat{H}$	REM	$\widehat{D}$	$\widehat{P}$
D	15170	22387	D	5560	0
H	18353	65754	P	2568	6830

Table 9. Performance Evaluation Results of NREM-1 and NREM-2 using the KMC Technique

Performance Measure	NREM-1				NREM-2		
	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$
Accuracy	0.769	0.767	0.996	0.984	0.639	0.628	0.930
Precision	0.939	0.736	–	0.788	0.453	0.716	0.087
Sensitivity	0.299	0.993	0.00	0.747	0.477	0.688	0.105
Specificity	0.991	0.364	100.00	0.993	0.718	0.522	0.960
F-Measure	0.45	0.84	0.00	0.77	0.46	0.70	0.10

Table 10. Performance Evaluation Results of NREM-3 and REM using the KMC Technique

Performance Measure	NREM-3		REM	
	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$
Accuracy	0.665	0.665	0.828	0.828
Precision	0.453	0.746	1.000	0.727
Sensitivity	0.404	0.782	0.684	1.000
Specificity	0.782	0.404	1.000	0.684
F-Measure	0.43	0.76	0.81	0.84

One of the standard classification algorithm is the Support Vector Machines (SVMs) [9]. The SVM is selected to perform on the same collected clinical dataset. The performance evaluation results are shown in Tables 11 and 12.

Tables 13 and 14 show a comparison summary of F-Measures in all of the selected techniques. The best classifying techniques are marked according to the sleep stages and the specified sleep disorder classes.

Table 11. Performance Evaluation Results of NREM-1 and NREM-2 using the SVM Technique

Performance Measure	NREM-1				NREM-2		
	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$
Accuracy	0.816	0.811	0.999	0.992	0.875	0.873	0.996
Precision	0.956	0.775	1.000	0.991	0.899	0.857	0.995
Sensitivity	0.447	0.992	0.893	0.764	0.697	0.961	0.897
Specificity	0.990	0.487	1.000	0.999	0.962	0.719	0.999
F-Measure	0.61	0.87	0.94	0.86	0.79	0.91	0.94

Table 12. Performance Evaluation Results of NREM-3 and REM using the SVM Technique

Performance Measure	NREM-3		REM	
	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$
Accuracy	0.825	0.825	0.992	0.992
Precision	0.951	0.804	0.979	0.999
Sensitivity	0.458	0.989	0.999	0.987
Specificity	0.989	0.458	0.987	0.999
F-Measure	0.62	0.89	0.99	0.99

Table 13. A Comparison Summary of F-Measures in NREM-1 and NREM-2

F-Measure	NREM-1				NREM-2		
	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$
CART	0.30	0.54	0.00	0.00	0.75	0.86	0.00
KMC	0.45	0.84	0.00	0.77	0.46	0.70	0.10
SVM	0.61*	0.87*	0.94*	0.86*	0.79*	0.91*	0.94*

Table 14. A Comparison Summary of F-Measures in NREM-3 and REM

F-Measure	NREM-3		REM	
	D-$\widehat{D}$	H-$\widehat{H}$	D-$\widehat{D}$	H-$\widehat{H}$
CART	1.00*	1.00*	1.00*	1.00*
KMC	0.43	0.76	0.81	0.84
SVM	0.62	0.89	0.99	0.99

5 Discussions and Implications

According to the classification results, the CART technique classification results are superior in the NREM-3 and the REM. There are a number of variables that have been selected as tree construction variables. For example, the SaO2 and Pulse variables are significantly nominated in all of the sleep stages, according to Table 2. However, there are a number of additional variables such as Abdomen that also engages into two of the

sleep stages, the NREM-2 and the NREM-3. Specially, there is only a variable, O2.A1, which is in the category of EEG. The SVM achieves the best classification performance in the NREM-1 and the NREM-2. The KMC technique has performed with lower accuracy. One of the possible explanations is the KMC technique utilises all of the variables for the classification without pre-feature selection stages.

Implications in terms of medical diagnosis, the distinction sleep disorder diagnostic is able to be performed by the PSG test in modern hospitals. On the other hand, the PSG test is not applicable in some areas due to its high costs. Our initial research discovers only a small set of variables that are main classifiers in the CART technique. The determined nominated variables may be pieces of evidences to classify various types of sleep disorders. Specifically, a traditional ECG and an oximeter may be able to use for overnight sleep tests. Additionally, other machine learning technique may be selected sequentially in order to construct a hybrid machine learning technique.

6 Conclusions

The sleep disorders are hidden symptoms that can provoke potential NCDs. The PSG test is recently the best accurate diagnostic tool per se in order to detect the sleep disorders. However, due to its high costs and availabilities, some researchers are seeking for alternative diagnostic tools and techniques that are able to partially or fully substitute the ordinary PSG test. Researchers attempt to select some properties from other measurements such as ECG, EEG, and EKG. Alternatively, a combination of signals may also be considered. The machine learning techniques have devoted as key mechanisms to classify sleep disorders. Most of the studies investigated only moderate to severe cases of sleep apnea. Based on our research works, other less severe sleep disorder subjects, which may lead to severe sleep apnea disorders, have been thoroughly studied. The selected data set is an original extracted dataset from the PSG test, which has not been altered or rescaled. The SVM achieves the best classification results in NREM-1 and NREM-2. The CART performs superior in NREM-3 and REM. The utmost goal is to minimise the number of nominated variables. The traditional ECG and the oximeter may be able to use for overnight sleep tests. It benefits rural hospitals or medical centres to achieve acceptable sleep disorder diagnosis results with standard medical devices. In terms of sleep disorder classifications, hybrid machine learning techniques will be investigated.

References

1. Centers for Disease Control and Prevention Information. http://www.cdc.gov/
2. Moser, D., Anderer, P., Gruber, G., Paraptics, S., Loretz, E., Boeck, M., Kloesch, G., Heller, E., Schmidt, A., Danker-Hopfe, H., Saletu, B., Zeitlhofer, J., Dorffner, G.: Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. In: Sleep, vol.32, pp. 139–149 (2009)

3. Ruehland, W., Rochford, P., O'Donoghue, F., Pierce, R., Singh, P., Thornton, A.: The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. Sleep **32**, 150–157 (2009)
4. Kocak, O., Bayrak, T., Erdamar, A., Ozparlak, L., Telatar, Z., Erogul, O.: Automated detection and classification of sleep apnea types using electrocardiogram (ECG) and electroencephalogram (EEG) features. In: Advances in Electrocardiograms – Clinical Applications, pp. 211–230 (2012)
5. Azarbarzin, A.Z.: Snoring sounds' statistical characteristics depend on anthropometric parameters. J. Biomed. Sci. Eng. **5**, 245–254 (2012)
6. Aksahin, M.F., Aydin, S., Firat, H., Erogul, O., Ardic, S.: Classification of sleep apnea types using EEG synchronization criteria. In: The 15th National Biomedical Engineering Meeting (BİYOMUT), pp. 1–4 (2010)
7. Khandoker, A.H., Palaniswami, M., Karmarkar, C.K.: Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings. IEEE Trans. Inf. Technol. Biomed. **13**, 37–48 (2009)
8. Liu, D., Pang, Z., Lloyd, S.: A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and EEG. IEEE Trans. Neural Netw. **19**, 308–318 (2008)
9. Hoffstin, V.: Apnea and snoring: state of the art and future direction. Acta Otorhinolaryngol Belg **56**, 205–236 (2002)
10. Berry, R.B.: Fundamentals of Sleep Medicine. Elsevier Saunders, Philadelphia (2012)
11. Wongsirichot, T., Iad-ua, N., Wibulkit, J.: A snoring sound analysis application using *k*-mean clustering method on mobile devices. In: Springer Series Advances in Intelligent Systems and Computing (2015)
12. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Waltham, MA (2012)
13. Costa, E.P., Lorena, A.C., Carvalho, A.C., Freitas, A.A.: A review of performance evaluation measures for hierarchical classifiers. In: Evaluation Methods for Machine Learning II, AAAI Press (2007)

A CLASSIFICATION OF SLEEP DISORDERS WITH OPTIMAL FEATURES USING MACHINE LEARNING TECHNIQUES

Thakerng Wongsirichot*, Anantaporn Hanskunatai

Department of Computer Science, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, 10520, Thailand

ABSTRACT:

Background: Sleep disorders become one of the early warnings of potential Non-Communicable Diseases (NCDs). Polysomnography (PSG) or sleep test is a formal method to diagnose sleep disorders. However, the PSG is limited in many hospitals due to its high costs. It also requires various sensors attached to a patient, which may cause inconvenience. Moreover, trained sleep specialists are required to interpret the gigantic PSG data. Researchers attempt to identify sleep disorders using alternative techniques.

Method: This study proposed an alternative technique for sleep-related syndrome and sleep disorder classification with optimal features. Patient PSG datasets were retrieved from a hospital in the south of Thailand. In the data preprocessing stage, the datasets were analyzed and normalized using feature extraction and selection mechanisms. Optimal feature selection using the average information gain values was evaluated with the 10-fold cross validation. Four Machine Learning (ML) techniques, *k*MC, *k*NN, SVM and MLP, were used in our experiments. The selected ML techniques have been performed and evaluated with the 10-fold cross validation in data preprocessing and model construction phases.

Results: The *k*NN achieved the highest overall classification results. The optimal features with *k*NN (*opf-k*NN) was proposed. The selected features were PULSE, SAO2, CANR and CHEST. With the selected optimal features, only the ordinary oxygen oximeter and the ECG machine were required. Overall classification result of the *opf-k*NN achieved at 95.17%±3.91.

Conclusion: Although the PSG is the formal sleep disorder diagnosis, alternative diagnostic techniques are beneficial especially to patients. Our study proposed the *opf-k*NN technique to classify sleep disorders with two concerns, the limited access to high-priced medical equipment and patient comfortability. Finally, sleep specialists also obtain benefits in optimizing bio-signal interpretations with only four optimal features.

Keywords: Sleep disorders; Polysomnography; Machine learning; Classification model

DOI: 10.14456/jhr.2017.26 *Received: July 2016; Accepted: November 2016*

INTRODUCTION

A number of people living with Non-Communicable Diseases (NCDs) or chronic diseases rose up to 133 million by the end of 2013 in the United States especially hypertension and cardiovascular diseases [1]. The WHO reported that 17.5 million people die of the cardiovascular diseases or 31% of all worldwide deaths in 2012 [2]. In Thailand, there were more than 400,000 NCD deaths and 29% of those were below the age of 60 in 2010 [3]. One of the early sign symptoms of the NCD is sleep disorders. Sleep disorders are found in both human and animals. There are a number of sleep disorder categories including parasomnias, dyssomnias, and circadian rhythm [4]. One common type of dyssomnias is Obstructive Sleep Apnea (OSA). The OSA is caused by the obstructions of the

* *Correspondence to: Thakerng Wongsirichot*
E-mail: thakerng.w@gmail.com

Cite this article as: Wongsirichot T, Hanskunatai A. A classification of sleep disorders with optimal features using machine learning techniques. J Health Res. 2017; 31(3): 209-17. DOI: 10.14456/jhr.2017.26

upper airway and series of repetitive pauses of breathing (apnea) during sleep. The American Academy of Sleep Medicine (AASM) stated the definition of the OSA when there are five or more detected apnea events per hour during sleep. The detections can be conducted in partial or full night sleep [4, 5]. Other kinds of early warning sleep-related syndromes are hypoventilation, hypoxemic, hypopnea syndromes, Restless Leg Syndrome (RLS), etc. Since 1930s, researchers attempted to describe hidden implications of sleep. The first full night Electroencephalogram (EEG) recording in human was performed [6]. In 1957, two phases of sleep were identified, the Non-Rapid Eye Movement (NREM) and the Rapid Eye Movement (REM).

Rechtschaffen and Kales (R&K) introduced a new scoring manual since 1968. Practically, the NREM stage consisted of four sub sleep stages and only one stage is the REM stage [7]. The most recent sleep stage classification standard is defined by the AASM. Specifically, the NREM consists of N-1, N-2 and N-3 sub-stages and there is only one stage in the REM. Sleep specialists interpret the sleep stages by reading bio-signals in the EEG. The N-1 or NREM-1 stage is a transition step from wakefulness and sleep. Slower heart rates are found in this stage with normal breathing parameters. A person can be easily awaked from external stimulus in the N-1 such as high volume noise, increasing of temperature, etc. Theta wave patterns are found in the EEG. The N-2 or NREM-2 stage is defined as a starting point of actual sleep. The overall muscle activity substantially decreases. A person gains less conscious awareness to external stimulus in the N-2 stage. Theta wave, sleep spindles and K-complexes wave patterns are presented in the EEG. The N-3 or NREM-3 stage is called a deep sleep stage. External stimulus have trivial effect on a person in the N-3 stage. Delta wave and a few sleep spindles are presented in the EEG. The R or REM stage occurs in latter half of sleep cycle before gaining consciousness. A person breathes more rapid with apparently eye movements in the REM stage. A mixed frequency of brain waves mostly low amplitude waves can be found in the EEG. In the sleep disorder diagnosis, Polysomnography (PSG) was used to classify sleep stages. Sleep stages occur in cycles and repeat during sleep. The PSG or sleep test records bio-signals during sleep. The bio-signals are collected by a full range of sensors attached to a patient's body. The sensors include a combination of Electrocardiographic (ECG), Electroencephalographic (EEG), Electrooculographic (EOG), and Electromyographic (EMG) [6, 8, 9].

RELATED WORK

Researchers in related fields proposed techniques that can possibly detect sleep disorders more efficiently. Most of the proposed techniques were constructed based on a combination of mathematical theories and machine learning techniques. A wide range of researches were conducted with different aspects. The identification of sleep stages is used to classify sleep disorders. A formal sleep stage classification framework was proposed for separating sleep stages automatically. It included a feature selection process and a classification process using MLP and flexible decision rules. The results showed 82% accuracy in the deep and paradoxical sleep [10]. A study of the effects of apnea duration in infants on changes of heart rates and oxygen desaturation levels was conducted. 236 apnea epochs were collected from multichannel recordings. The result showed both heart rates and oxygen desaturation levels were significantly related to the apnea durations [11]. An application of ambulatory ECG to identify OSA was studied. Time and frequency domains of Heart Rate Variability (HRV) analysis techniques had been engaged. The results reached 81.25% sensitivity, 46.81% specificity and 64.21% positive predictive value [12]. A Neural Network (NN) system was used to validate the oxygen desaturation levels from oximeters to predict OSA. The overall classification result was promisingly at 93.30% accuracy [13]. A screening OSA test on mobile phones was initiated. A group of researchers built a mobile application so called SleepAp that records Photoplethysmogram (PPG) signals via mobile oximeters. The classification using the Support Vector Machine (SVM) reached 88.40% accuracy in the first experiment and 92.30% accuracy in the second experiment [14].

MATERIALS AND METHODS

Polysomnography (PSG) or sleep test is a formal sleep disorder diagnosis recording a full range of signals from brain, heart, muscle, etc. [15]. A set of full PSG multichannel recordings have been collected from full night sleep (8-12 hours) of patients. The recordings were extracted from the Sleepscan VISION software [16] at the Songklanagarin Hospital, Hat Yai, Songkhla,

Table 1 The PSG signal attributes

Category	Features	Description
Electroencephalogram (EEG)	C3-A2, C4-A1	Monopolar EEG in a position C3-A2, C4-A1
	F3-A2, F4-A1	Bipolar EEG in a position F3-A2, F4-A1
	O1-A2, O2-A1	Monopolar EEG in a position O1-A2, O2-A1
Body and muscle movement	LOC, ROC	Left / right outer canthus
	CHEST	Chest movement
	ABDO	ABDO movement
	CHIN	Chin movement
	LAT, RAT	Left / right anterior tibialis
Electrocardiography (ECG)	ECG	Electrocardiography
	PULSE	Pulse
Thoracic respiratory efforts	CANR	Nasal airflow
	FLOW	Mouth airflow
	SAO2	Oxygen desaturation
	SNR	Amplitude of snoring sounds

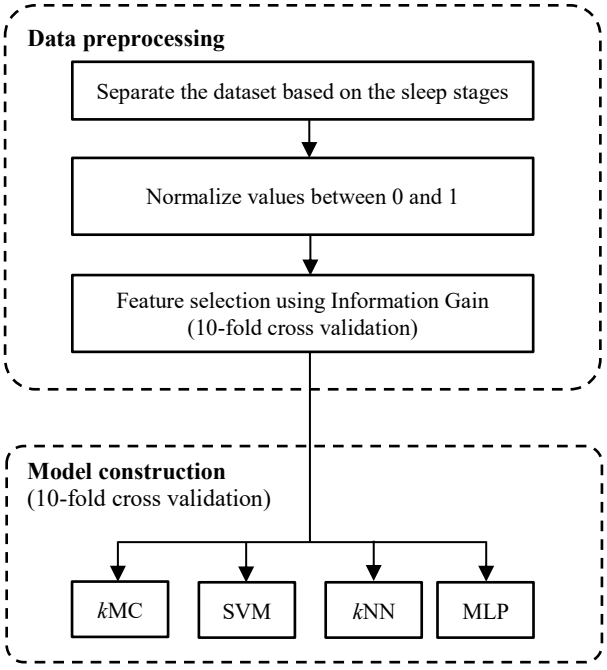


Figure 1 The proposed classification model

Thailand. This study was approved by the hospital's director with explicit attentions to the patient confidentiality.

The collected dataset, Table 1, was initially retrieved from five subjects. The subjects were two males and three females (age 39±11.2 years, BMI 27.75±4.5 kg/m^2, AHI 8.2±2.0). It contained 440,593 records. The dataset consisted of four classes of sleep disorders, including Oxygen Desaturation (D), Hypopnea (H), Isolated Limb Movement (I) and Periodic Limb Movement (P) that were sequentially indicated in every 30-second epochs by professional sleep technicians. The objective of this study was to classify sleep-related syndromes and sleep disorders with optimal features using machine learning techniques. In order to achieve a common classification model, the retrieved datasets, which were retrieved from different patients, were combined as one, without subjective concerns. The proposed classification model consists of two main steps, the data preprocessing and the model construction as shown in Figure 1.

Data preprocessing

The dataset was divided into sleep stages, N-1, N-2, N-3, and R. Each record represented in a form of discrete signal that was a sequence of values in specific time series. Three consecutive data records

Time-stamp	C3-A2	C4-A1	…	Sleep disorder
10478.75	-0.1874	7.4874	…	D
10478.75	-1.7853	7.7363	…	D
10478.75	-3.7690	7.3395	…	D
…	…	…	…	…
16032.01	-18.449	0.9919	…	P
16032.01	-17.258	6.9429	…	P
16032.01	-17.227	7.2478	…	P

Figure 2 An example of the original dataset

Table 2 Information gain values of features in sleep stages

Features	Information gain				
	N-1	N-2	N-3	R	Avg
PULSE	0.297	0.300	0.066	0.675	0.335
SAO2	0.197	0.192	0.087	0.594	0.268
CANR	0.050	0.042	0.015	0.402	0.127
CHEST	0.043	0.040	0.018	0.266	0.092
CHIN	0.030	0.004	0.016	0.272	0.081
ABDO	0.062	0.052	0.020	0.156	0.073
ROC	0.170	0.027	0.037	0.054	0.072
FLOW	0.040	0.030	0.016	0.142	0.057
F3-A2	0. 150	0.018	0.008	0.015	0.048
ECG	0.140	0.010	0.002	0.020	0.043
O1-A2	0.029	0.043	0.030	0.044	0.037
LOC	0.025	0.027	0.031	0.047	0.033
O2-A1	0.006	0.038	0.025	0.055	0.031
C3-A2	0.017	0.034	0.017	0.013	0.020
C4-A1	0.007	0.011	0.010	0.015	0.014
LAT	0.005	0.011	0.004	0.026	0.012
F4-A1	0.007	0.007	0.010	0.019	0.011
RAT	0.004	0.007	0.001	0.008	0.005
SNR	0.003	0.006	0.000	0.000	0.002

with the same timestamp represented one second in time, as shown in Figure 2.

The attribute original values in the dataset were in different scales. Therefore, all of the numeric attributes were normalized between 0 and 1. The 10-fold cross validation method is used to evaluate models especially in classification and prediction problems. In practice, an original dataset is randomly partitioned into ten portions (P_1, P_2, …, P_{10}). A portion (P_i) of the ten portions is selected as a test set. Explicitly, every data portion is tested exactly once and becomes a training set nine times. The training sets are used to construct the classification model and the test sets are used to measure the performance of the classification models. The process is iteratively running until each of the portions has been tested [17]. The Information Gain (IG) is a heuristic function that quantifies abilities of features or attributes in classifying data. A calculation of IG values is to identify high IG features that are used as a set of selected features in a classification model [17-19]. In each of the sleep stages, IG was used to statistically measure each of the features with the 10-fold cross validation method. Table 2 showed the feature selection result using IG. Specifically, in a specific sleep stage, each feature was evaluated with the 10-fold cross validation method. Finally, each feature has ten IGs from the 10-fold cross validation method. The average value of the ten IGs was calculated. Moreover, the average IGs of a feature in all of sleep stages are averaged into one averaged IG values as shown in "avg" column. The average of IGs were used as a ranker of the attributes.

Model construction

The model construction was designed to compare classification performances of four selected ML techniques including *k*-Mean Clustering (*k*MC), *k*-Nearest Neighbor (*k*NN),

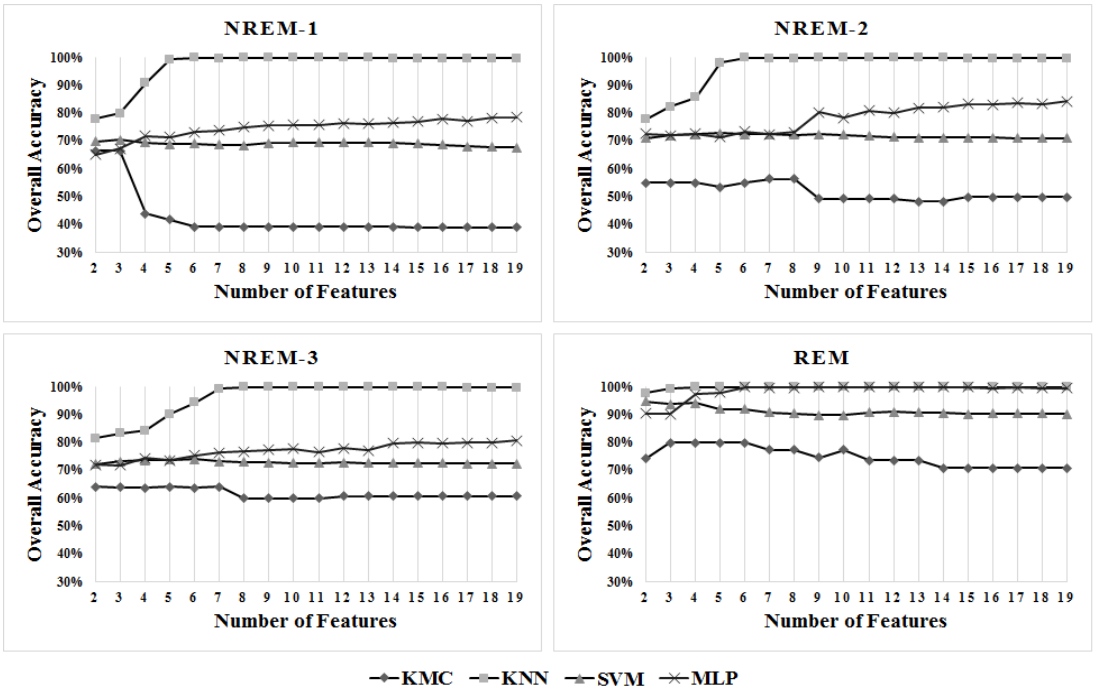


Figure 3 Overall classification accuracy results of the selected ML techniques

Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). The 10-fold cross validation method was used to evaluate the performances of the selected ML techniques. The *k*MC is an unsupervised machine learning technique. It assigns each of the instances in a dataset into exactly one cluster by measuring its similarity without consideration of a predefined class. For classification problems, the classified instance is assigned to a cluster in which the distance between the instance and the cluster is minimum. The *k*NN is a supervised machine learning technique. The number of nearest neighbors (*k*) is set. The *k* nearest neighbors are selected by calculating the distances between all instances in the training set and the classified instance. If the *k* nearest neighbors is set more than one, the instance is classified into a class by a majority vote of the *k* nearest neighbors [17]. The SVM is a supervised machine learning technique. It classifies data into classes by using a hyperplane. The hyperplane is an optimal separator, which calculated from a training set. The original SVM is based on a binary classification category [17]. The MLP is a supervised machine learning technique in the Artificial Neuron Network (ANN). It stimulates the concept of human brain that comprises of a number of neurons or nodes connected together. Each of the neurons contain a computational function so called activation function. It adjusts weights of nodes to improve classification or prediction performance according to a specified learning rate [17].

In our study, there are a number of predetermined conditions and settings. In the *k*MC, the number of *k* clusters was set based on the number of sleep disorder classes in each sleep stages in our experiment. Additionally, the *k*MC was solely used in one of our previous research works to identify the sleep disorders [20, 21]. In the *k*NN, the *k* value was set to 1. In the SVM, the kernel function was selected to the Radial Basis Function (RBF) with the Euclidean Distance function. The RBF was recommended in similar research works [14]. In multi-class problems, more than one SVM binary classifiers are used. The ensemble of SVM classifiers plays an important role with error correcting mechanisms to improve overall classification accuracy. In the MLP, different learning rates were recommended and tested. The suitable learning rate was set to 0.03.

EXPERIMENTAL RESULTS

Figure 3 represented the overall classification accuracy results of the selected ML techniques in each of the sleep stages. The graphs can be classified into three common patterns. Firstly, the accuracies increase when the number of features increase. Secondly, the accuracies substantially decrease

Table 3 A summary of overall classification results of the machine learning (ML) techniques

Sleep stage	ML technique	Overall accuracy	Weighted sensitivity	Weighted F measure	Weighted ROC
N-1	*k*MC	42.79%±0.09	0.50±0.06	0.52±0.03	0.55±0.01
	*k*NN	96.85%±0.07	0.97±0.07	0.97±0.07	0.98±0.05
	SVM	69.13%±0.01	0.69±0.01	0.62±0.01	0.59±0.01
	MLP	74.39%±0.04	0.74±0.04	0.70±0.05	0.81±0.04
N-2	*k*MC	51.62%±0.03	0.52±0.03	0.55±0.02	0.59±0.00
	*k*NN	96.81%±0.07	0.97±0.07	0.97±0.07	0.98±0.04
	SVM	71.78%±0.01	0.72±0.01	0.68±0.01	0.65±0.01
	MLP	78.31%±0.05	0.78±0.05	0.78±0.05	0.77±0.06
N-3	*k*MC	61.64%±0.02	0.62±0.02	0.62±0.01	0.58±0.00
	*k*NN	96.24%±0.07	0.96±0.07	0.96±0.07	0.97±0.06
	SVM	72.90%±0.01	0.73±0.01	0.60±0.01	0.57±0.01
	MLP	77.09%±0.03	0.77±0.03	0.46±0.06	0.74±0.04
R	*k*MC	77.95%±0.02	0.79±0.02	0.79±0.02	0.79±0.04
	*k*NN	99.64%±0.01	1.00±0.00	1.00±0.00	1.00±0.00
	SVM	92.28%±0.02	0.92±0.02	0.92±0.02	0.94±0.01
	MLP	96.95%±0.04	0.97±0.03	0.97±0.03	0.97±0.03

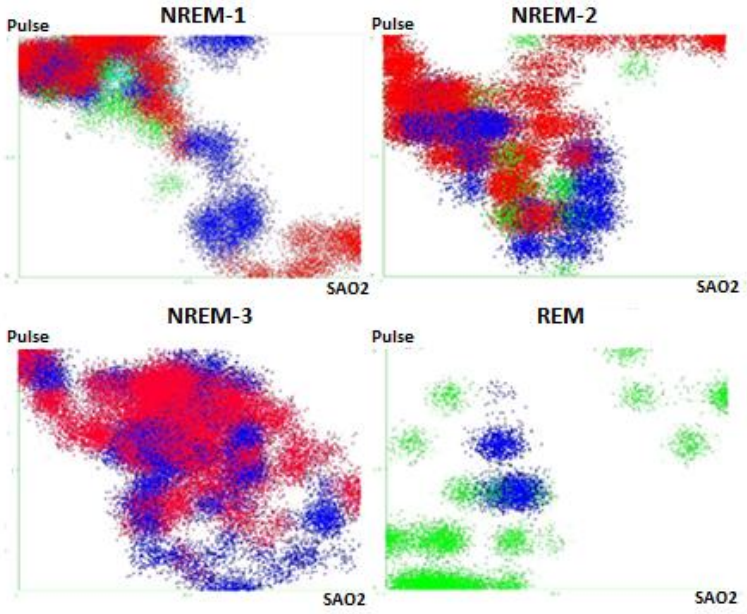


Figure 4 Data visualization of PULSE and SAO2 in all of the sleep stages

when the number of features increase. Finally, the number of features have no effect the accuracies. The *k*NN is classified into the first category. It reveals that the increasing of the number of features reflects the accuracies. On the other hand, the *k*MC classification performance results decrease when the number of features increase, which is categorized in the second case.

Table 3 summarized overall classification results of the selected ML techniques. The measurements were in a form of $\bar{X}$ ± SD. There are four main measurements presented including overall accuracy, weighted sensitivity, weighted specificity and weighted ROC. The calculation of the measurements were conducted based on the number of selected features and their ranked orders. Specifically, the first calculation was conducted with 19 features based on the ranked order as shown in Table 2. The second calculation omitted the lowest IG feature, SNR so there were 18 selected for the calculation. The calculation process continued until the last two highest IG features, Pulse and SAO2, were used in the calculation. All of the measurements were weighted due to the differences

Table 4 A comparison of classification results of the *k*NN(*2f*), *k*NN(*5f*), *opf-k*NN, *best-k*NN

Sleep stage	ML technique	Overall accuracy	Weighted sensitivity	Weighted F measure	Weighted ROC
N-1	*k*NN*(2f)*	77.91%±0.03	0.78±0.03	0.32±0.01	0.77±0.01
	*k*NN*(5f)*	90.82%±0.02	0.91±0.03	0.91±0.03	0.91±0.02
	*opf-k*NN	92.67%±0.05	0.93±0.03	0.93±0.02	0.93±0.02
	*best-k*NN	99.98%±0.02	1.00±0.00	1.00±0.00	1.00±0.00
N-2	*k*NN*(2f)*	77.81%±0.04	0.78±0.03	0.23±0.04	0.78±0.04
	*k*NN*(5f)*	85.73%±0.04	0.86±0.04	0.86±0.02	0.86±0.03
	*opf-k*NN	96.64%±0.03	0.97±0.03	0.97±0.03	0.97±0.03
	*best-k*NN	99.99%±0.01	1.00±0.00	1.00±0.00	1.00±0.01
N-3	*k*NN*(2f)*	81.51%±0.04	0.82±0.03	0.36±0.04	0.80±0.04
	*k*NN*(5f)*	84.33%±0.03	0.84±0.04	0.84±0.04	0.82±0.05
	*opf-k*NN	91.40%±0.03	0.91±0.03	0.91±0.03	0.91±0.03
	*best-k*NN	99.96%±0.02	0.99±0.01	1.00±0.01	0.92±0.02
R	*k*NN*(2f)*	97.75%±0.03	0.98±0.03	0.98±0.02	0.98±0.03
	*k*NN*(5f)*	99.98%±0.02	1.000±0.01	1.000±0.01	1.000±0.01
	*opf-k*NN	99.98%±0.02	1.000±0.01	1.000±0.01	1.000±0.01
	*best-k*NN	100.00%±0.00	1.000±0.00	1.000±0.00	1.000±0.00
AVERAGE	*k*NN*(2f)*	83.75%±9.49	0.84±0.10	0.23±0.16	0.83±0.10
	*k*NN*(5f)*	90.22%±7.08	0.90±0.07	0.90±0.07	0.90±0.08
	*opf-k*NN	95.17%±3.91	0.95±0.04	0.95±0.04	0.95±0.04
	*best-k*NN	99.98%+0.02	0.99±0.01	1.00±0.00	0.98±0.04

of sleep disorder classes.

The best classification results were *k*NN, MLP, SVM and *k*MC, respectively. The overall accuracies of the *k*MC were not promising. The accuracies of the *k*MC reduced dramatically at specific number of features. Theoretically, the *k*MC partitions observations into clusters with the nearest mean. An instance is classified into a cluster based on the calculation of distance between the instance and a nearest cluster. Overlapping between instances in different classes lead to higher misclassification rates. On the other hand, the *k*NN classifies clusters based on the *k* nearest points. Multiple *k* nearest points can be used to classify an instance into a cluster [17] so that it is a possible reason that the *k*NN outperformed the *k*MC in our study. Theoretically, the SVM and the MLP are more robust in classification problems. Finally, the SVM and the MLP seems to have steady classification performance trends. The overall accuracies of the SVM and the MLP are not strongly affected by the alteration of number of features.

Figure 4 showed a data visualization of two axes, namely PULSE and SAO2, which obtained the top two highest average IGs in Table 2. It revealed the *k*MC classified incompetently in the dataset that contained a lot of overlapping instances especially in N-1, N-2 and N-3. In this case, the data distribution in REM was relatively round and less overlapping so it performed reasonably better as shown in Table 3. In addition, the *k*NN confirmed the classification performance was relatively high in the REM. Specifically, there were two D (blue) groups that were located separately to each other and the P (green) group in the REM. The data points were uniformly distributed so that the *k*NN was a suitable classification technique, especially in REM [22].

The first two highest average IG features were PULSE and SAO2, 0.335 and 0.268 in Table 2. The rest obtained the average IGs less than 0.13. Therefore, the features, PULSE and SAO2, were selected as the first two nominated features. The PULSE and SAO2 are a subset of the PPG, which are measured by an ordinary oximeter [14]. Other selected features after PULSE and SAO2 were CANR, CHEST and CHIN, respectively. Technically, CANR and CHEST are monitored using an ordinary ECG. CHIN is recorded by the EMG sensor attached to a patient's chin, which is less patient comfortability. Therefore, the selected optimal features were PULSE, SAO2, CANR and CHEST with considerations of less equipment and patient comfortability. A further experiment was used to measure the *k*NN classification performances with a different number of features. Firstly, the *k*NN(*2f*) consisted of two features, PULSE and SAO2. Secondly, the *k*NN(*5f*) selected only top five ranked features by the average IG values. Thirdly, the *opf-k*NN, the proposed model,

used only four features including PULSE, SAO2, CANR and CHEST. Finally, the *best-k*NN achieved the maximum classification results regardless of a number of selected features. The number of selected features in the *best-k*NN were 12, 12, 16 and 5 in N-1, N-2, N-3, and R, respectively.

Table 4 showed a comparison of overall classification results of *k*NN(*2f*), *k*NN(*5f*), *opf-k*NN and *best-k*NN. The performance measures were evaluated using the 10-fold cross validation method. In average, the *best-k*NN obtained the highest average accuracy at 99.98%+0.02. The second highest was the *opf-k*NN at 95.17%±3.91. The *k*NN*(5f)* was ranked as the third position from the best one with five selected features followed by the *k*NN*(2f)*. Our main intention targeted to the *opf-k*NN that employed appropriate features. Therefore, a paired-sample *t* test was conducted to compare the classification accuracies between each of the *k*NN algorithms, which employed different numbers of features. Significance level was defined as a *P* value of less than 0.05. Firstly, there was no statistically significant difference between the mean of correctly classified classes ($p = 0.18$) of the *best-k*NN and the *opf-k*NN. Secondly, there was statistically significant difference between the mean of correctly classified classes ($p = 0.0085$). According to the paired-sample *t* test results, the *opf-k*NN achieved the overall classification accuracy was comparable to the *best-k*NN with only four optimal features. Moreover, the *opf-k*NN also statistically outperformed the *k*NN(*2f*). The paired-sample *t* tests were conducted based on all of the gathered datasets in order to confirm the generalization of the classification model.

DISCUSSION

In our study, there were a number of factors required for considerations. The detection of sleep disorders normally requires a combination of oxygen oximeter, ECG, EEG, and PSG. The PSG alone includes all of the essential signal sensors for sleep disorder diagnoses. However, the PSG is expensive due to the setup cost and its hardware and software. Some patients feel unconformable during the sleep test due to the attaching sensors. More economical medical equipment is also available separately that can be used for the diagnosis. The main purpose was to technically reduce a number of features to the sleep-related syndrome and sleep disorder classifications. There were two main reasons to obtain the optimal feature set, a reduction of high-priced medical equipment and an improvement of patient comfortability during the diagnosis. Specifically, the statistical test showed that the *best-k*NN and the *opf-k*NN is no statistically significant difference in terms of classification performances. However, the *best-k*NN required the maximum number of features. The proposed *opf-k*NN, required only 4 features including PULSE, SAO2, CANR and CHEST in which only required the ordinary oxygen oximeter and the ECG machine. It can be concluded that the *opf-k*NN overcame the *best-k*NN in terms of costs and patient comfortability with no statistically significant difference. The discovery of the optimal features and the *opf-k*NN technique can assist hospitals or medical centers that are not equipped with the PSG. The proposed model can be used as a screening test. In addition, sleep technicians gain advantages to optimize bio-signal interpretations with only four optimal features. This initial study aids in determining minimum features required for sleep disorder classification. In the case of developing countries, our study can benefit small or medium hospitals that are not equipped with the PSG. There are several limitations to our study. Our subjects were diagnosed with positive OSA symptoms, which do not mimic the diversity of entire population. Another limitation is the selected PSG machine and setting used in the hospital was only from one manufacturer. A comparison to other PSG machines is an advantage.

CONCLUSION

Sleep disorders become one of the early warning signs of the NCDs. A number of researchers applied ML techniques in order to classify the sleep disorders, especially the OSA. The objective of our study was to classify sleep-related syndromes and sleep disorders with optimal features. The average IG was used to rank the optimal features. The optimal features included PULSE, SAO2, CANR and CHEST. A comparison of different ML techniques was thoroughly conducted. The highest classification accuracy was the *k*NN followed by MLP, SVM and *k*MC, respectively. An advanced step was performed to compare the *k*NN with different number of features. Our proposed model, *opf-k*NN, adapted the *k*NN to analyze with the optimal features. The overall classification results were at 95.17%±3.91. The *opf-k*NN achieved the overall classification accuracy is statistically comparable to the *best-k*NN, the best classifier with

maximum number of features. This study had no intention to discover the technique that achieved the highest classification scores but also concern about the minimum use of high-priced medical equipment and maintaining patient comfortability. In common, this study was a systematic investigation to identify effective features for archiving classification goals. The investigation targeted into in-depth data analyses. This study attempted to discover a generalized sleep disorder classification model. The population size could be one of the concerns. However, this study initially reviewed more than 400,000 records of bio-signal data from subjects. The dataset covered both males and females and widely spreading of age 39±11.2. This study benefits in two aspects. Firstly, it can be designed as a screening test with minimum medical equipment and sensors used. Secondly, sleep technicians can apply this technique to optimize bio-signal interpretations with only four optimal features.

REFERENCES

1. Centers of Disease Control and Prevention [CDC]. [cited 2015 September 7]. Available from: http://www.cdc.gov/
2. World Health Organization [WHO]. Cardiovascular disease, Fact sheet N°317; 2009 [cited 2015 September 7. Available from: http://www.who.int/mediacentre/factsheets/fs317/en/index.html
3. Institute of Medical Research and Technology Assessment. Literature review: the current situation and care model of non-communicable diseases. Nonthaburi: Ministry of Public Health, Institute of Medical Research and Technology Assessment; 2014.
4. American Academy of Sleep Medicine [AASM]. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep. 1999 Aug; 22(5): 667-89.
5. Azarbarzin A, Moussavi Z. Snoring sounds' statistical characteristics depend on anthropometric parameters. Journal of Biomedical Science and Engineering. 2012; 5(5): 245-54. doi: 10.4236/jbise.2012.55031
6. Jafari B, Mohsenin V. Polysomnography. Clinics in Chest Medicine. 2010; 31(2): 287-97. doi: 10.1016/j.ccm.2010.02.005
7. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects. Washington DC: Washington Public Health Service, US Government Printing Office; 1968.
8. Haba-Rubio J, Krieger J. Evaluation instruments for sleep disorders: a brief history of polysomnography and sleep medicine. In: Chiang RPY, Kang SC, editors. Introduction to modern sleep technology. Netherlands: Springer; 2012. p.19-31.
9. Berry R. Fundamentals of sleep medicine. Philadelphia: Elsevier Saunders; 2012.
10. Chapotot F, Becq G. Automated sleep-wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules. International Journal of Adaptive Control and Signal Processing. 2010; 24:409-23. doi: 10.1002/acs.1147
11. Carbone T, Marrero LC, Weiss J, Hiatt M, Hegyi T. Heart rate and oxygen saturation correlates of infant apnea. J Perinatol. 1999 Jan; 19(1): 44-7.
12. Sun J, Li X, Guo J, Han F, Zhang H. Identification of obstructive sleep apnea syndrome by ambulatory electrocardiography: clinical evaluation of time-domain and frequency-domain analyses of heart rate variability in Chinese patients. Cell Biochem Biophys. 2011 Apr; 59(3): 165-70. doi: 10.1007/s12013-010-9128-6
13. Almazaydeh L, Faezipour M, Elleithy K. A neural network system for detection of obstructive sleep apnea through SpO2 signal features. International Journal of Advanced Computer Science and Applications. 2012; 3(5):7-11.
14. Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, et al., SleepAp: an automated obstructive sleep apnoea screening application for smartphones. Computing in Cardiology. 2013; 40: 257-60.
15. Robertson B, Marshall B, Carno MA. Polysomnography for the sleep technologist: instrumentation, monitoring, and related procedures. 1st ed. St. Louis: Elsevier Mosby; 2014.
16. Natus Medical Incorporated. Natus medical incorporated. 2012 [cited 2015 June 12]. Available from: http://www.natus.com/
17. Han J, Kamber M, Pei J. Data mining: concepts and techniques. Waltham: Morgan Kaufmann; 2011.
18. Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. In: ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning; 1997. San Francisco, CA, USA: Morgan Kaufmann; 1997. p. 412-420.
19. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explorations. 2009; 11: 10-8.
20. Wongsirichot T, Iad-ua N, Wibulkit J. A snoring sound analysis application using K-Mean clustering method on mobile devices. In: Kacprzyk J, editor. Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015. Wroclaw, Poland: Springer; 2015. p. 789-796.
21. Wongsirichot T, Hanskunatai A. A comparative investigation of PSG signal patterns to classify sleep disorders using machine learning techniques. In: Intelligent computing theories and methodologies. Fuzhou, China: Springer International Publishing Switzerland; 2015. p. 510-21.
22. Domeniconi C, Peng J, Gunopulos D. Locally adaptive metric nearest-neighbor classification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002 Sep; 24(9): 1281-5. doi: 10.1109/Tpami.2002.1033219

Biomedical Engineering: Applications, Basis and Communications, Vol. 30, No. 6 (2018) 1850041 (13 pages)
DOI: 10.4015/S1016237218500412

World Scientific
www.worldscientific.com

A MULTI-LAYER HYBRID MACHINE LEARNING MODEL FOR AUTOMATIC SLEEP STAGE CLASSIFICATION

Thakerng Wongsirichot* and Anantaporn Hanskunatai
Advanced Artificial Intelligence Research Laboratory
Department of Computer Science, Faculty of Science
King Mongkut's Institute of Technology Ladkrabang (KMITL)
Bangkok 10520, Thailand

Accepted 5 June 2018
Published 1 October 2018

ABSTRACT

Sleep Stage Classification (SSC) is a standard process in the Polysomnography (PSG) for studying sleep patterns and events. The SSC provides sleep stage information of a patient throughout an entire sleep test. A physician uses results from SSCs to diagnose sleep disorder symptoms. However, the SSC data processing is time-consuming and requires trained sleep technicians to complete the task. Over the years, researchers attempted to find alternative methods, which are known as Automatic Sleep Stage Classification (ASSC), to perform the task faster and more efficiently. Proposed ASSC techniques usually derived from existing statistical methods and machine learning (ML) techniques. The objective of this study is to develop a new hybrid ASSC technique, Multi-Layer Hybrid Machine Learning Model (MLHM), for classifying sleep stages. The MLHM blends two baseline ML techniques, Decision Tree (DT) and Support Vector Machine (SVM). It operates on a newly developed multi-layer architecture. The multi-layer architecture consists of three layers for classifying W, R and $N1$, $N2$, $N3$ in different epoch lengths. Our experiment design compares MLHM and baseline ML techniques and other research works. The dataset used in this study was derived from the ISRUC-Sleep database comprising of 100 subjects. The classification performances were thoroughly reviewed using the hold-out and the 10-fold cross-validation method in both subject-specific and subject-independent classifications. The MLHM achieved a certain satisfactory classification results. It gained 0.694 ± 0.22 of accuracy (AUC $= 0.822 \pm 0.31$) in subject-specific classification and 0.942 ± 0.02 of accuracy (AUC $= 0.920 \pm 0.17$) in subject-independent classification. The pros and cons of the MLHM with the multi-layer architecture were thoroughly discussed. The effect of class imbalance was rationally discussed towards the classification results.

Keywords: Automatic sleep stage classification; Hybrid machine learning; Multi-layer classification model; Data mining.

INTRODUCTION

Sleep is a fundamental resting state of humans. During sleep, humans are physically unconscious, however many activities are functioning. Studies have been comprehensively conducted in order to understand human sleep processes for various purposes. One of the vital purposes is to identify sleep disorders which can cause further health problems. In order to classify sleep stages, a patient requires visiting a Polysomnography (PSG), or a sleep test. In PSG, bio-signals are digitally gathered via non-invasive sensors and electrodes. Trained sleep technicians thoroughly analyze the collected

*Corresponding author: Thakerng Wongsirichot, Department of Computer Science, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, 10520, Thailand. Tel: +66-846414784; E-mail: thakerng.w@gmail.com

bio-signal data to classify sleep stages and patterns. The process is called Sleep Stage Classification (SSC). The SSC is a gold standard method to study sleep stages. Results from SSCs are used for diagnosing sleep disorders such as sleep apnea, insomnia, parasomnia, etc. The first proposal of SSC was introduced by Rechtschaffen and Kales (R&K) in 1968. R&K proposed two groups of sleep stages, Rapid Eye Movement (REM) and Non-Rapid Eye Movement (NREM). The NREM consisted of four sleep stages, $N1$, $N2$, $N3$ and $N4$, and there was only one stage in the REM. Since 2007, the American Academy of Sleep Medicine (AASM) redefined the sleep stages and renounced a new definition of the SSC. The AASM standard consists of three sleep stages in NREM ($N1$, $N2$, $N3$) and only one sleep stage in REM. $N1$ and $N2$ are the light sleep stages. $N3$ and R are the deep sleep stages. Currently, the SSC complies with the AASM standard.[1–3]

A PSG or a sleep test is a formal medical process in gathering vital bio-signals for identifying sleep disorders. A PSG consists of Electroencephalogram (EEG), Electrooculography (EOG), Electromyography (EMG) and Electrocardiogram (ECG). EEG is a non-invasive testing technique that derives from a number of electrodes, which are attached to a patients scalp, for analyzing brain activities.[4] During an EEG test, any pairs of electrodes measure differences between them via differential amplifiers. Figure 1 shows the EEG electrode placement based on the AASM 10–20 system with the recommended and backup derivations marked.[4] For example, F4–A1 measures the signal differences between F4 and A1 electrodes. As per the AASM standard for the SSC, the recommended derivations are F4–A1, C4–A1, O2–A1, and the corresponding backup derivations are F3–A2, C3–A2, O1–A2, respectively.[4,5]

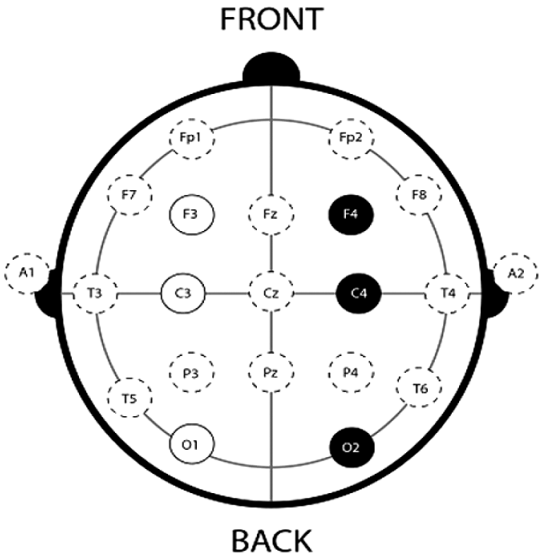


Fig. 1 EEG electrode placement.

In general, a trained sleep technician visually localizes specific waveforms of EEG signals in order to classify sleep stages, Wakefulness (W), $N1$, $N2$, $N3$ and R. To discovering alternative ASSC, multiple EEG signals have been used in various research in both times and frequency domains.[6–13] There are six common types of waveforms in the EEG. Alpha rhythm is an oscillation waveform with a frequency between 8 and 13 Hz that can be found in W, $N1$, and R. Sleep spindle is presented in the spindle-shaped waveform with frequency between 11 and 16 Hz. The sleep spindle is commonly found in $N2$ and $N3$. K complex is an extremely high amplitude waveform that continuously shows both high negative and positive wave amplitudes. The K complex can be detected in $N2$ and $N3$. The lengths of both of the sleep spindle and the K complex are more than 0.5 s. Vertex sharp wave is a sharp waveform with a very short duration (< 500 ms) that is presented in $N1$. Slow wave activity is a high amplitude waveform with a frequency between 0.5 and 2 Hz that is visualized in $N2$ and $N3$. The slow wave activity is last between 0.5 and 2 s. Saw-tooth wave is a triangular or saw-like waveform with a frequency between 2 and 6 Hz. The saw-tooth is found in R. Figure 2 shows samples of 10 s of C4–A1. In the $N1$, the vertex sharp waves were presented in three marked positions. In the $N2$, there was almost 5 s of the sleep spindle with a rising amplitude similar to K complex wave presented. In the $N3$, two marked very high amplitude of K complex waves were presented. In the R, a series of the saw-tooth wave was presented in triangular waveforms. Finally, in the W, alpha rhythm was marked at the beginning and the rest was steady waves.[4]

EOG records differences of the voltage of eye movement sensors. AASM scoring manual recommends placing two electrodes, one above the Right Outer Canthus (ROC) and another one below the Left Outer Canthus (LOC). The recommended derivations are LOC-A2 and ROC-A1 and the alternative derivations are LOC-Fz and ROC-Fz. There are three eye movement patterns in EOG that associate with the SSC. Firstly, the eye blink is a vertical eye movement that can be detected in W with a frequency between 0.5 and 2 Hz. Secondly, the slow eye movement is a sinusoidal eye movement that is only presented in W and $N1$. The duration of the slow eye movement is more than 500 ms. Thirdly, the REM is an extremely high and sharp amplitude waveform that is found in R. The duration of REM wave is usually less than 500 ms.[1,5]

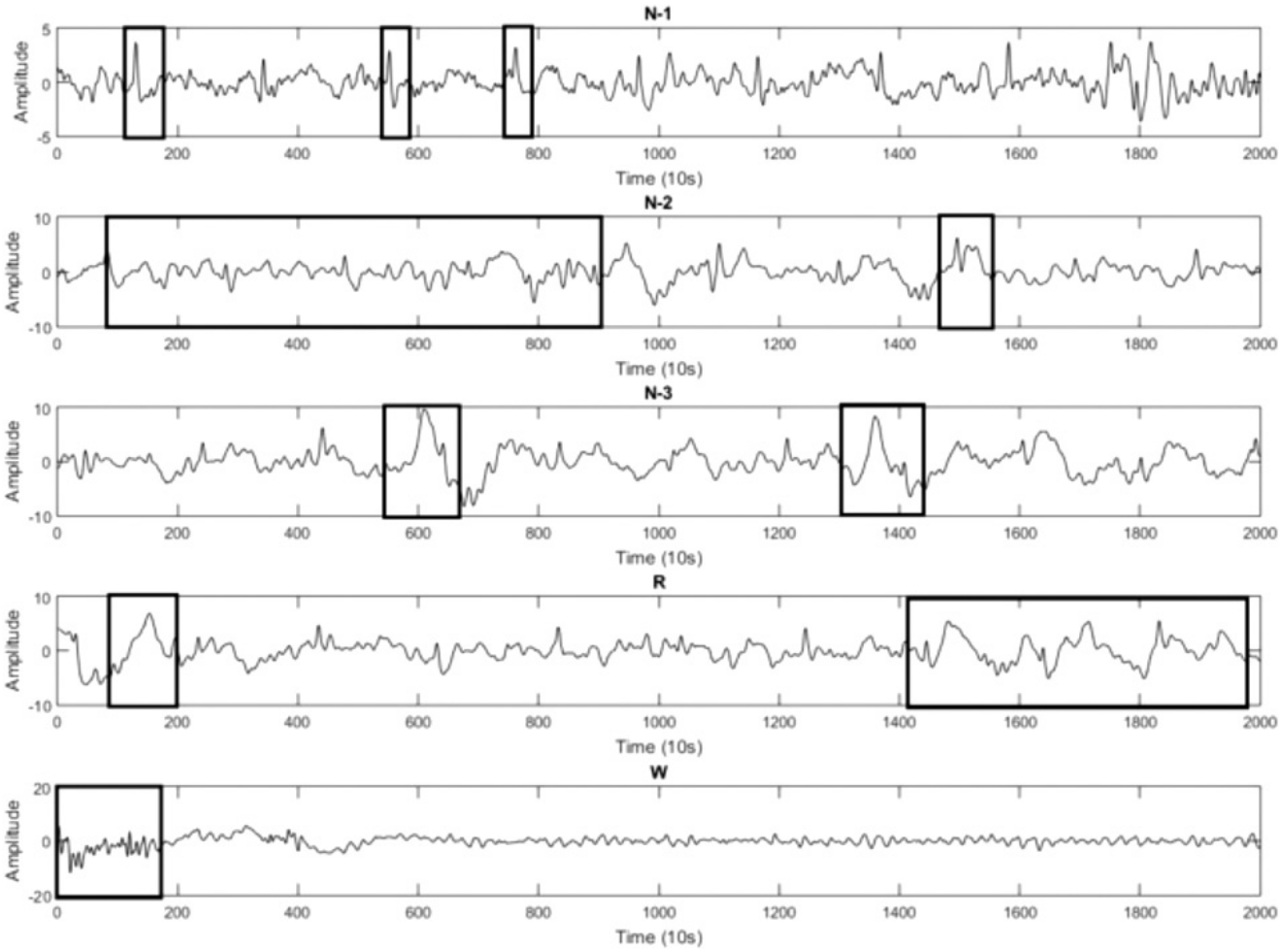


Fig. 2 C4–A1 signal in sleep stages (10 s epoch length).

Figure 3 shows samples of 10 s of EOG signal from the LOC-A2 electrode in sleep stages. In $N1$, the slow eye movements are presented for more than 7 s. In $N2$, there is absolutely no existence of slow eye movement. In $N3$, there are no specific EOG patterns. In R, there are REM waveforms presented. Finally, in the W, the slow eye movement is marked.

EMG records electrical activities of skeletal muscles by placing multiple sensors on specific muscles. The sensors can be placed in many body areas such as chin, legs, hands and other skeletal muscles. Chin EMG is a common signal derived from electrodes usually placed under a patients chin. According to the AASM standard, there are three electrodes that are used for chin EMG. The first electrode is placed in the center of the chin. The second and third electrodes are placed under the chin on the left and right-hand side, respectively. The chin EMG is mainly used for detecting the R stage. It distinguishes between the NREM and REM stages by comparing the reduction of amplitude of the chin EEG waves. Figure 4 shows a sample of chin EMG signal. The amplitude of the chin EMG was dropped dramatically during the REM comparing with the rest of the signal, which was in the NREM.

In ECG test, a patient is placed with a number of electrodes on the limb. It mainly records electrical activities of a heart. Electrodes and sensors of pulse, respiratory effort, airflow are usually placed in a patients chest. ECG signals have been widely used in classifying sleep stages. For instance, Heart Rate Variability (HRV) in both frequency and time domains are widely used in the SSC. HRV is derived from calculating a number of values from both ECG and respiratory effort.[1,4,14]

SSC is a time-consuming process. It requires many working hours to visually perform on an individual subject by trained sleep technicians. Researchers proposed methods to perform Automatic Sleep Stage Classification (ASSC). The methods derived from a combination of existing mathematical, statistical and

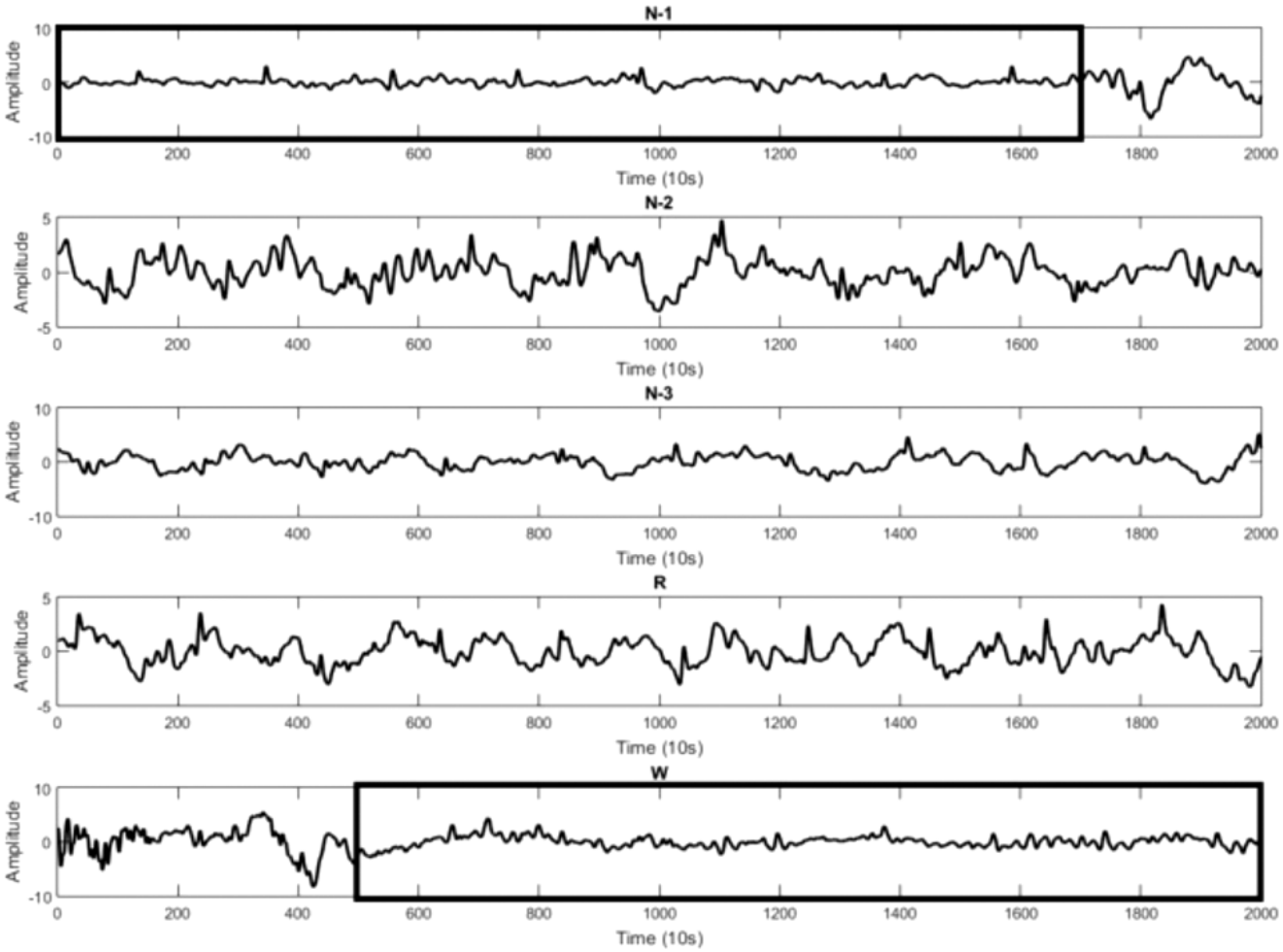


Fig. 3 LOC–A2 signal in sleep stages (10 s epoch length).

computer science related theories and techniques. One of the most selected computer science related techniques over the years is in the machine learning (ML) discipline. In general, the ASSC process consists of two main parts. Firstly, the signal processing and feature selection process is initially processed raw data from an original PSG to make it applicable for a developed ASSC. Feature selection techniques can be categorized into two main groups, full PSG features and partial PSG features. Partial PSG features are selected by using statistical or ML techniques. A variety of signal processing techniques has been applied in the SSC problems. The techniques include wavelet transform,[6,7] spectral entropy,[8,15] and statistical values such as median, standard deviation, arithmetic mean, skewness and kurtosis.[9,10,16] A comparison of uses of phase space and power spectral approaches for wake-sleep identification were studied.[11]

Secondly, the SSC process is performed using various techniques. The classification model can be very simple as one selected model or very complex by combining various existing well-known techniques, which is called a hybrid technique. The SVM-DT hybrid approach

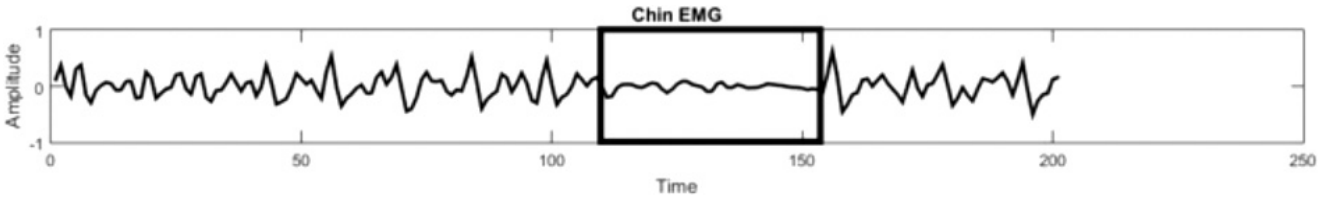


Fig. 4 Chin EMG signal wave forms.

was embedded Support Vector Machine (SVM) into Decision Tree (DT) as a DT prepruning method. It claimed to improve classification performance. Five datasets were selected for the experiment. The classification performances were slightly higher when compared with the baseline DT and SVM.[17] The Dendrogram-based SVM (DSVM) framework was introduced. The DSVM used the multi-class SVM classification based on a DT approach. Feature selections had been conducted. The k-fold cross-validation had been performed to evaluate the DSVM framework. The result was acceptable when comparing with sleep technician manual scoring.[18] A multi-class SVM had been used in another research work. The main objective was to automatically classify sleep stages from EEG signals. The work achieved 92% accuracy.[12] Sirvan Supervised Method for Sleep Staging (SSM4S) was proposed using SVM with selected features. The SSM4S was evaluated with different selected features from EEG, EOG and EMG. The SSM4S used two SVM models for classifying sleep–wake stages and classifying multi-class sleep stage.[19] A comparative investigation of baseline ML techniques, k-Mean Clustering (KMC), CART and SVM, in classifying sleep disorders was conducted. SVM was the best classifier for $N1$ and $N2$. The CART outperformed in $N3$ and R. A finding that could be further investigated was only two selected features, Oxygen Saturation and Pulse, were used in the CART.[20] The use of Oxygen Saturation and Pulse together with HRV has been evidenced in various research works.[21–23] Some distinctive research works were conducted by using snoring sound analysis for detecting sleep apnea. A group of researchers collected data from 28 snorers, 22 with Obstructive Sleep Apnea (OSA) and six without OSA, to analyze their snoring sound patterns. Anthropometric parameters and statistical characters of each of the snorers were analyzed. Two features, Median Bifrequency (MBF) and Projected MBF (PMBF) were used together with the anthropometric parameters such as age, gender, BMI, etc.[24] With the evolution of mobile devices, a mobile application for snoring sound analysis was proposed for sleep apnea classification. KMC and SMOTE techniques were utilized in this study. The classification result reached up to 80.10%.[25] SleepAp was an automated OSA screening application in a smartphone. Selected features included audio, actigraphy, photoplethysmography (PPG) and demographics were used with SVM classifier. A total number of subjects were 856 (735 for training and 121 for testing). It achieved up to 92.2% accuracy for classifying subjects with moderate or severe OSA.[26]

MATERIALS AND METHODS

This study develops a new hybrid ASSC technique, the Multi-Layer Hybrid ML Model (MLHM), for classifying sleep stages. It combines two baseline ML techniques, DT and SVM, in a newly developed multi-layer architecture. The motivation of this study is to mimic manual SSC process taken by sleep technicians and combine with two selected baseline ML techniques into a hybrid model. In addition, one of the distinctions of our hybrid model compares with other existing hybrid models by adding multiple classification layers. The MLHM consists of two main steps, the Data Preprocessing and the Hybrid Classification Model as presented in Fig. 5. In this study, the MATLAB software package was used.

Dataset

The ISRUC-Sleep Subgroup I dataset used in this study was derived from the ISRUC-Sleep database. It is made publicly available online.[27] It was collected at the Sleep Medicine Center of the Hospital of Coimbra University (CHUC) in Portugal during 2009–2013. The dataset has been used in a number of research works for automatic SSC problems.[13,19,28–30] It contained bio-signals from

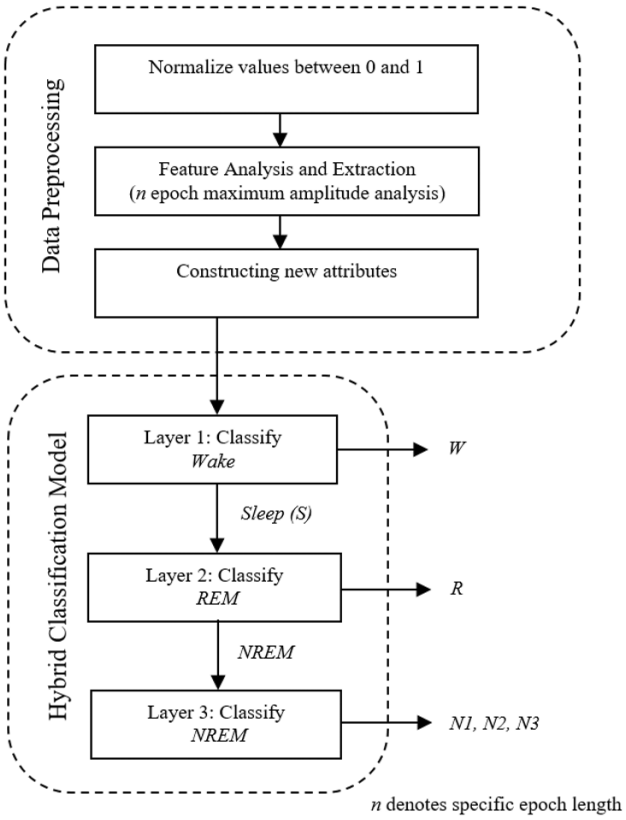


Fig. 5 Multi-layer hybrid ML model (MLHM).

Table 1. ISRUC-Sleep Subgroup 1 Dataset Structure.

Attribute	Description
EEG	
C3–A2	Monopolar EEG in a position C3–A2
C4–A1	Monopolar EEG in a position C4–A1
F3–A2	Bipolar EEG in a position F3–A2
F4–A1	Bipolar EEG in a position F4–A1
O1–A2	Monopolar EEG in a position O1–A2
O2–A1	Monopolar EEG in a position O2–A1
EOG	
LOC–A2	Left Outer Canthus
ROC–A1	Right Outer Canthus
EMG	
Chin EMG	Placed between the chin and lower lip
Leg-1 EMG	Left leg movement
Leg-2 EMG	Right leg movement

100 human adults, including both healthy subjects and subjects with sleep disorders. The subjects comprised of 53 male, 42 female and five unspecified sex, age 51 ± 16 years, height 1.35 ± 6.63 m, weight 65.09 ± 34.97 kg, BMI $23.53 \pm 12.83\,\mathrm{kg/m^2}$. Each of the subjects attended approximately 8–9 h full night PSG test. Technically, all of the recorded signals were sampled at 200 MHz and segmented into 30 s epoch length according to the AASM standard. This dataset contained bio-signals from 11 electrodes in EEG, EOG and EMG as shown in Table 1.[17,27] There is a sleep technician which has completely interpreted sleep stages for each of the specific recorded signals.

Data Preprocessing

Data preprocessing stage consists of three sub-processes including normalization, feature extraction and selection and constructing. In this dataset, each of the 30 s epoch length records contained a series of 6000 continuing values. In order to omit the scale differences between the signals, the statistical normalization (between 0 and 1) was applied to each of the signals independently. The scale differences may lead to possibilities of dominant attributes. The normalization formula is shown in Eq. (1).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \tag{1}$$

where x is an original value of a particular signal group. x' is a normalized value of the original x.

Generally, a number of feature analysis and extraction techniques were used in ASSC research problems for identifying important signal pattern characteristics. Signal patterns can be differentiated by its wave patterns, frequencies and amplitudes. A single or a combination of feature analysis and extraction technique was proposed and used in many works.[31] In this study, all of the features in Table 1 were taken into account. Two feature analysis and extraction techniques were selected, the Fast Fourier Transform (FFT) and the Maximum Amplitude Analysis (MAA), respectively. The FFT relies on the Fourier Transform Analysis concept. It computationally transforms continuous signals from their original domain, in this study the time domain, to the frequency domain representation.[32] The signal transformation shows the frequency of the original signals that can assist in the interpretations. The FFT has been widely used in many research works .[9,13,19,20,27–30] In addition, the outputs from the FFT were located the maximum amplitudes using the MAA in each of the epoch lengths, 5, 10, 15 and 30 s. Figure 6 shows C4–A1 signals after the FFT process with the maximum amplitude marked in two consecutive 30 s epoch length. The MAA is used to extract a specific feature that assist the AASC.

According to the current AASM standard and guidelines, there are five sleep stages including W, R, $N1$, $N2$ and $N3$. In the manual SSC process, sleep technicians visually locate three distinct groups of bio-signal patterns, sleep–wake, NREM–REM and $N1$–$N2$–$N3$.[3,4] In order to mimic the manual SSC process, two additional attributes, SW and NR, have been constructed for the MLHM without reducing the total number of features. The SW attribute is designed for the Layer I classification. It consists of Sleep(S) and Wake (W). The S includes all of NREM and REM. The NR attribute is designed to distinguish NREM and REM in the Layer II classification. It includes the N for all

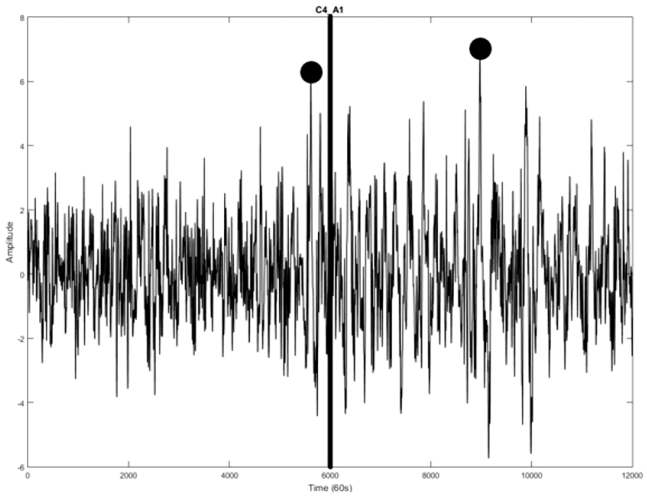


Fig. 6 Maximum amplitude analysis.

Table 2. A Sample of *N*3 Signal in Specified Epoch Lengths.

Ep (s)	C3–A2	C4–A1	F3–A2	F4–A1	O1–A2	O2–A1	LOC–A2	LOC–A1	Chin EMG	Leg-1 EMG	Leg-2 EMG	*SS*	*SW*	*NR*
30	0.913	0.966	0.988	0.972	0.977	0.998	0.980	0.971	0.985	0.951	0.987	*N*3	*S*	*N*
15	0.575	0.515	0.511	0.513	0.563	0.511	0.520	0.509	0.496	0.508	0.510	*N*3	*S*	*N*
	0.913	0.966	0.988	0.972	0.977	0.998	0.980	0.965	0.985	0.951	0.987			
	0.575	0.515	0.511	0.513	0.563	0.511	0.518	0.509	0.496	0.508	0.510			
10	0.885	0.966	0.965	0.956	0.977	0.983	0.980	0.965	0.985	0.951	0.987	*N*3	*S*	*N*
	0.913	0.958	0.988	0.972	0.972	0.998	0.968	0.971	0.984	0.951	0.987			
	0.557	0.515	0.511	0.513	0.563	0.480	0.518	0.509	0.457	0.485	0.510			
	0.575	0.494	0.491	0.486	0.503	0.511	0.494	0.507	0.496	0.508	0.480			
5	0.557	0.484	0.511	0.484	0.487	0.468	0.520	0.507	0.478	0.488	0.500	*N*3	*S*	*N*
	0.885	0.966	0.965	0.956	0.977	0.983	0.980	0.965	0.985	0.951	0.987			
	0.899	0.915	0.988	0.945	0.887	0.998	0.920	0.889	0.924	0.951	0.933			
	0.913	0.958	0.968	0.972	0.972	0.978	0.968	0.971	0.984	0.947	0.987			

NREM ($N1, N2, N3$) and R for REM. Specifically, the W is classified in the Layer I, therefore, the W is not included in the Layer II. In addition, the R is classified in the Layer II and not classified in the Layer III of the MLHM. On the completion of the data preprocessing process, data records consist of data values from all of the 11 attributes in 30, 15, 10 and 5 s epoch length. The data records are mapped with original sleep stage, *SW* and *NR*. Table 2 depicts a sample of N3 in different epoch length after the data preprocessing.

Hybrid Classification Model

In this study, the hybrid classification model is constructed using two ML techniques, DT and SVM. The DT is a non-parametric supervised ML method. It can manage both categorical and numerical data. A training dataset is required in the learning process of the DT. From the learning process, it identifies a set of decision rules for classifying target variables. The DT has been claimed to be one of the simplest techniques for understandings and interpretations.[32] The SVM is widely used in classification problems. It classifies target variables-based selected learning algorithms. In common, a kernel function is selected to perform a non-linear classification by mapping the training dataset into high dimensional feature spaces. In this study, the Radial Basis Function (RBF) was selected as our kernel function.[32]

In addition, a new multi-layer architecture was designed together with the hybrid classification model. The multi-layer architecture consisted of three layers. The MLHM is designed to perform binary classifications in the Layers I and II. For the Layer III, the multi-class classification is used. According to Fig. 5, the Layer I classified the S and W as designated in the *SW* attribute. The output of the Layer I was the W. The S instances were passed to the Layer II. The Layer II was designed to differentiate between NREM and REM instances as assigned in the NR attribute. The output of the Layer II was the REM. The NREM instances were passed to the Layer III. The third layer classified three classes, $N1$, $N2$ and $N3$, simultaneously.

Table 3. Ranked Classifiers and Epoch Lengths.

Layer	ML(epoch)	ACC	Sens	Spec	*F*	AUC
*L*1	DT(5ep)*	0.633	0.7688	0.2286	0.7026	0.7701
(W)	DT(10ep)	0.628	0.7639	0.2352	0.6997	0.7644
	DT(30ep)	0.626	0.7577	0.2400	0.6937	0.7589
	SVM(5ep)	0.444	0.4647	0.5367	0.4303	0.4640
	SVM(30ep)	0.295	0.1800	0.8200	0.1494	0.1800
*L*2	SVM(5ep)*	0.793	0.9038	0.0951	0.8394	0.9044
(R)	DT(5ep)	0.764	0.8695	0.1084	0.8184	0.8806
	DT(10ep)	0.748	0.8465	0.1300	0.8078	0.8583
	DT(15ep)	0.745	0.8418	0.1394	0.8024	0.8512
	DT(30ep)	0.740	0.8349	0.1459	0.7972	0.8445
*L*3	DT(15ep)*	0.681	0.8124	0.1866	0.7719	0.8124
($N1$, $N2$, $N3$)	DT(5ep)	0.680	0.8101	0.1917	0.7708	0.8100
	DT(30ep)	0.676	0.8032	0.1981	0.7675	0.8031
	DT(10ep)	0.676	0.7978	0.2032	0.7729	0.7977
	SVM(5ep)	0.656	0.7982	0.2038	0.5313	0.7981

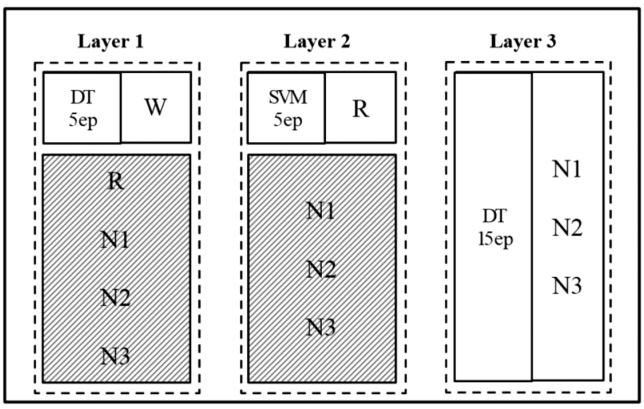


Fig. 7 The formation of hybrid classification model in the MLHM.

Table 4. Subject-Independent Classification Results.

Sleep Stage	Perform. Measure	MLHM	SS4MS	DT	SVM
All sleep stages	Accuracy	0.942 ± 0.02	0.902 ± 0.08	0.932 ± 0.04	0.943 ± 0.03
	Sensitivity	0.891 ± 0.27	0.752 ± 0.16	0.803 ± 0.12	0.873 ± 0.11
	Specificity	0.954 ± 0.27	0.942 ± 0.05	0.953 ± 0.03	0.962 ± 0.03
	*F*measure	0.852 ± 0.14	NA	0.784 ± 0.12	0.813 ± 0.15
	AUC	0.920 ± 0.17*	0.845 ± 0.11	0.875 ± 0.15	0.915 ± 0.30
W	Accuracy	0.953 ± 0.01	0.941 ± 0.05	0.945 ± 0.04	0.949 ± 0.04
	Sensitivity	0.891 ± 0.16	0.882 ± 0.11	0.801 ± 0.17	0.874 ± 0.17
	Specificity	0.951 ± 0.32	0.951 ± 0.05	0.953 ± 0.03	0.972 ± 0.02
	*F*measure	0.843 ± 0.07	NA	0.764 ± 0.18	0.792 ± 0.23
	AUC	0.920 ± 0.26*	0.915 ± 0.04	0.875 ± 0.20	0.920 ± 0.28*
R	Accuracy	0.961 ± 0.01	0.894 ± 0.02	0.947 ± 0.04	0.964 ± 0.03
	Sensitivity	0.892 ± 0.15	0.822 ± 0.20	0.821 ± 0.09	0.884 ± 0.07
	Specificity	0.973 ± 0.04	0.971 ± 0.04	0.964 ± 0.03	0.982 ± 0.03
	*F*measure	0.882 ± 0.06	NA	0.793 ± 0.09	0.863 ± 0.12
	AUC	0.930 ± 0.28*	0.895 ± 0.15	0.890 ± 0.21	0.930 ± 0.22*
*N*1	Accuracy	0.920 ± 0.02	0.881 ± 0.05	0.925 ± 0.03	0.938 ± 0.03
	Sensitivity	0.912 ± 0.23	0.391 ± 0.17	0.702 ± 0.13	0.882 ± 0.13
	Specificity	0.933 ± 0.08	0.952 ± 0.04	0.963 ± 0.02	0.942 ± 0.03
	*F*measure	0.721 ± 0.45	NA	0.704 ± 0.13	0.663 ± 0.21
	AUC	0.920 ± 0.17*	0.670	0.830 ± 0.22	0.910 ± 0.21
*N*2	Accuracy	0.900 ± 0.03	0.855 ± 0.01	0.903 ± 0.04	0.906 ± 0.05
	Sensitivity	0.853 ± 0.12	0.80 ± 0.15	0.82 ± 0.09	0.84 ± 0.08
	Specificity	0.952 ± 0.06	0.88 ± 0.07	0.93 ± 0.04	0.95 ± 0.03
	*F*measure	0.892 ± 0.07	NA	0.83 ± 0.10	0.85 ± 0.09
	AUC	0.900 ± 0.29*	0.840 ± 0.16	0.875 ± 0.18	0.895 ± 0.25
*N*3	Accuracy	0.950 ± 0.04	0.942 ± 0.08	0.947 ± 0.03	0.959 ± 0.03
	Sensitivity	0.921 ± 0.68	0.842 ± 0.17	0.854 ± 0.01	0.893 ± 0.09
	Specificity	0.964 ± 0.84	0.971 ± 0.04	0.973 ± 0.02	0.984 ± 0.02
	*F*measure	0.923 ± 0.06	NA	0.854 ± 0.10	0.893 ± 0.09
	AUC	0.940 ± 0.30*	0.905 ± 0.05	0.910 ± 0.33	0.935 ± 0.22

Note: *The highest AUC.

Classification performance measures include accuracy, sensitivity, specificity, F measure and Area Under the Curve (AUC). As the 10-fold cross-validation method was conducted in this study, the AUC in each of the folds were averaged to measure the classification performance. The AUC is a quantitative representation of the ROC curve. It is one of the most selected performance measures in imbalance problems. The AUC can be calculated as the arithmetic means of TP and TN.[33] Eqs. (2)–(6) are the selected performance measures used in this study.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \tag{2}$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \tag{3}$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})}, \tag{4}$$

$$F \text{ Measure} = \frac{2\text{TP}}{(2\text{TP} + \text{FP} + \text{FN})}, \tag{5}$$

$$\text{AUC} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + FP}\right). \tag{6}$$

In order to construct an efficient ASSC model, the entire dataset was investigated. The subject-independent classification was conducted using the hold-out technique. The full dataset (all subjects) has been separated into 70:30. The first 70 was used as a training set with 10-fold cross-validation method. The remaining 30 was used as a test set. DT and SVM with different epoch lengths have been evaluated in each of the layers. For example, the classification of W with SVM(30ep) denoted the use of SVM with 30 s epoch length. The selection was performed based on the training set.

Table 3 depicts top five classification results of the training set in each designed layers descending by AUC. The Layer 1, which classified W, achieved the highest value of the AUC by DT with 5 s epoch length. Classifying R in the Layer 2 reached the maximum AUC value by SVM with 5 s epoch length. In the Layer 3, the multi-class classification of $N1$, $N2$ and $N3$ maximized

Table 5. Subject-Specific Classification Results.

Sleep Stage	Perform. Measure	MLHM	DT(5ep)	DT(10ep)	DT(15ep)	DT(30ep)	SVM(5ep)	SVM(10ep)	SVM(15ep)	SVM(30ep)
All	Acc	0.694 ± 0.22	0.688 ± 0.21	0.681 ± 0.21	0.677 ± 0.20	0.678 ± 0.22	0.631 ± 0.22	0.501 ± 0.26	0.502 ± 0.26	0.501 ± 0.26
sleep	Sens	0.822 ± 0.33	0.810 ± 0.32	0.802 ± 0.32	0.776 ± 0.32	0.800 ± 0.33	0.712 ± 0.32	0.518 ± 0.40	0.520 ± 0.40	0.518 ± 0.40
stages	Spec	0.802 ± 0.33	0.782 ± 0.31	0.752 ± 0.32	0.734 ± 0.35	0.736 ± 0.33	0.682 ± 0.32	0.524 ± 0.40	0.528 ± 0.40	0.520 ± 0.40
	F	0.748 ± 0.29	0.744 ± 0.28	0.736 ± 0.27	0.734 ± 0.28	0.736 ± 0.28	0.678 ± 0.28	0.458 ± 0.35	0.458 ± 0.35	0.458 ± 0.37
	AUC	0.822 ± 0.31**	0.813 ± 0.32*	0.798 ± 0.33*	0.780 ± 0.33*	0.798 ± 0.32*	0.711 ± 0.39*	0.519 ± 0.40*	0.521 ± 0.33*	0.520 ± 0.40*
W	Acc	0.633 ± 0.24	0.633 ± 0.24	0.633 ± 0.24	0.617 ± 0.24	0.625 ± 0.24	0.444 ± 0.28	0.300 ± 0.25	0.300 ± 0.25	0.295 ± 0.25
	Sens	0.772 ± 0.33	0.771 ± 0.33	0.763 ± 0.33	0.754 ± 0.35	0.762 ± 0.35	0.472 ± 0.42	0.181 ± 0.39	0.182 ± 0.39	0.182 ± 0.39
	Spec	0.751 ± 0.33	0.757 ± 0.33	0.722 ± 0.33	0.683 ± 0.36	0.688 ± 0.35	0.462 ± 0.42	0.248 ± 0.38	0.253 ± 0.38	0.244 ± 0.39
	F	0.714 ± 0.30	0.718 ± 0.30	0.695 ± 0.29	0.698 ± 0.30	0.691 ± 0.30	0.435 ± 0.40	0.154 ± 0.33	0.154 ± 0.33	0.157 ± 0.33
	AUC	0.769 ± 0.34**	0.769 ± 0.34	0.764 ± 0.34	0.746 ± 0.35	0.757 ± 0.35	0.465 ± 0.42*	0.180 ± 0.48*	0.180 ± 0.39*	0.180 ± 0.39*
R	Acc	0.793 ± 0.22	0764 ± 0.20	0.748 ± 0.21	0.745 ± 0.22	0.740 ± 0.22	0.741 ± 0.22	0.565 ± 0.36	0.565 ± 0.36	0.566 ± 0.36
	Sens	0.904 ± 0.29	0.854 ± 0.27	0.852 ± 0.28	0.843 ± 0.29	0.833 ± 0.29	0.802 ± 0.29	0.591 ± 0.48	0.592 ± 0.48	0.591 ± 0.49
	Spec	0.891 ± 0.29	0.822 ± 0.24	0.782 ± 0.29	0.764 ± 0.29	0.753 ± 0.29	0.882 ± 0.29	0.544 ± 0.49	0.544 ± 0.48	0.552 ± 0.49
	F	0.844 ± 0.26	0.828 ± 0.24	0.811 ± 0.24	0.801 ± 0.26	0.801 ± 0.26	0.845 ± 0.26	0.553 ± 0.45	0.554 ± 0.45	0.551 ± 0.45
	AUC	0.904 ± 0.31**	0.870 ± 0.29*	0.848 ± 0.29*	0.840 ± 0.30*	0.836 ± 0.31*	0.894 ± 0.31	0.592 ± 0.39*	0.592 ± 0.49*	0.593 ± 0.49*
$N1$	Acc	0.802 ± 0.22	0.800 ± 0.22	0.801 ± 0.22	0.772 ± 0.12	0.801 ± 0.22	0.790 ± 0.20	0.732 ± 0.29	0.732 ± 0.29	0.732 ± 0.30
	Sens	0.911 ± 0.28	0.911 ± 0.28	0.912 ± 0.28	0.802 ± 0.28	0.917 ± 0.29	0.891 ± 0.26	0.822 ± 0.39	0.824 ± 0.39	0.823 ± 0.39
	Spec	0.881 ± 0.28	0.863 ± 0.28	0.872 ± 0.27	0.772 ± 0.28	0.872 ± 0.28	0.852 ± 0.27	0.844 ± 0.39	0.846 ± 0.39	0.802 ± 0.39
	F	0.852 ± 0.26	0.858 ± 0.26	0.853 ± 0.25	0.855 ± 0.26	0.853 ± 0.26	0.851 ± 0.22	0.764 ± 0.36	0.764 ± 0.36	0.761 ± 0.36
	AUC	0.911 ± 0.28**	0.908 ± 0.28*	0.910 ± 0.27	0.811 ± 0.28*	0.910 ± 0.28	0.895 ± 0.26*	0.820 ± 0.37*	0.820 ± 0.39*	0.820 ± 0.39*
$N2$	Acc	0.566 ± 0.18	0.565 ± 0.18	0.558 ± 0.18	0.546 ± 0.16	0.566 ± 0.18	0.659 ± 0.10	0.609 ± 0.16	0.613 ± 0.16	0.609 ± 0.16
	Sens	0.722 ± 0.39	0.713 ± 0.39	0.694 ± 0.39	0.682 ± 0.39	0.711 ± 0.39	0.854 ± 0.13	0.833 ± 0.37	0.844 ± 0.36	0.833 ± 0.37
	Spec	0.702 ± 0.39	0.693 ± 0.39	0.645 ± 0.39	0.673 ± 0.39	0.682 ± 0.40	0.704 ± 0.13	0.805 ± 0.36	0.814 ± 0.37	0.802 ± 0.37
	F	0.604 ± 0.32	0.605 ± 0.32	0.594 ± 0.31	0.602 ± 0.32	0.611 ± 0.31	0.782 ± 0.08	0.675 ± 0.29	0.681 ± 0.29	0.676 ± 0.30
	AUC	0.719 ± 0.39	0.711 ± 0.39*	0.688 ± 0.39*	0.719 ± 0.39	0.714 ± 0.39	0.751 ± 0.31*	0.831 ± 0.38*	0.840 ± 0.37*,**	0.832 ± 0.36*
$N3$	Acc	0.674 ± 0.22	0.676 ± 0.22	0.667 ± 0.22	0.674 ± 0.24	0.662 ± 0.22	0.519 ± 0.31	0.305 ± 0.24	0.305 ± 0.24	0.306 ± 0.24
	Sens	0.815 ± 0.34	0.814 ± 0.33	0.802 ± 0.33	0.814 ± 0.31	0.796 ± 0.33	0.553 ± 0.49	0.172 ± 0.38	0.173 ± 0.38	0.174 ± 0.38
	Spec	0.791 ± 0.34	0.792 ± 0.33	0.755 ± 0.33	0.792 ± 0.42	0.705 ± 0.34	0.523 ± 0.49	0.203 ± 0.38	0.203 ± 0.38	0.214 ± 0.38
	F	0.741 ± 0.29	0.748 ± 0.28	0.744 ± 0.28	0.734 ± 0.28	0.734 ± 0.28	0.492 ± 0.43	0.163 ± 0.34	0.153 ± 0.33	0.161 ± 0.39
	AUC	0.807 ± 0.33**	0.807 ± 0.33**	0.795 ± 0.32*	0.807 ± 0.34**	0.785 ± 0.33*	0.548 ± 0.49*	0.173 ± 0.40*	0.172 ± 0.38*	0.174 ± 0.38*

Notes: *There is statistically significant difference between the mean of AUC of the MLHM and a corresponding classification model at p-value of 0.05.
**The highest AUC.

the AUC value by DT with 15 s epoch lengths, respectively. Therefore, the formation of hybrid classification model in the MLHM was constructed as shown in Fig. 7.

RESULTS

In this study, the classification performance of the MLHM in the subject-independent classification was compared with DT, SVM and SSM4S. The SSM4S was performed in 30 s epoch length.[19] Therefore, DT and SVM used the same configuration, the 30 s epoch length, in this comparison. Table 4 shows a comparison of the subject-independent classification results on the test set of DT, SVM, MLHM and SSM4S. The results are depicted in a form of $\bar{x} \pm S.D.$ In average, the MLHM outperformed other techniques in all of the sleep stages according to the AUC. The MLHM gained the highest AUC in $N1$, $N2$ and $N3$. The MLHM and SVM achieved equal AUC in W and R. The SVM gained slightly higher accuracy comparing with the MLHM, however, it achieved equal or lower AUC. The MLHM reached average accuracy up to 0.942 ± 0.02 (AUC $= 0.920 \pm 0.17$) and the SSM4S achieved average accuracy in all of the sleep stages at 0.902 ± 0.08 (AUC $= 0.845 \pm 0.11$).

The aforementioned subject-independent classification depicted general aspects of the MLHM model. The MLHM performed relatively well compared with DT, SVM and SSM4S. However, this study concerned on the possibility of general use of the MLHM in clinical tests. It was worth to investigate more specific in a subject-specific study or a specific study. The MLHM was applied to classify sleep stages in each of the individual subjects in the dataset. In addition, DT and SVM with specific epoch lengths were tested in the subject-specific study. The subject-specific classification results of the SSM4S were not presented in the original work.[29] Table 5 shows the subject-specific classification results of all sleep stages. There were nine classification models that were used to test on each subject individually. The models include DT, SVM with specific epoch lengths (5, 10, 15 and 30 s) and the MLHM. The average performance measures were calculated from all 100 subjects in a form of $\bar{x} \pm S.D.$ The MLHM achieved the highest accuracy and AUC, 0.694 ± 0.22 and 0.822 ± 0.31. The DT(5ep) was the second-to-best achieving 0.688 ± 0.21 of accuracy and 0.813 ± 0.32 of AUC. Specifically, the DT achieved the higher rank comparing with the SVM in each of the specified epoch.

Paired-sample t-tests were performed to statistically validate the classification performances between MLHM comparing and baseline ML techniques in the subject-specific classifications. The significance level was set at a p-value of 0.05.

Table 6. Imbalanced Ratios of Subject-Independent Classifications in Test Sets.

Perform. Measures	Layer 1 $W:S$	Layer 2 REM:NREM	Layer 3 $N1$:non-$N1$	Layer 3 $N2$:non-$N2$	Layer 3 $N3$:non-$N3$
IR	3.734	4.773	4.022	1.030	2.243
Acc (%)	0.953 ± 0.01	0.961 ± 0.01	0.920 ± 0.02	0.900 ± 0.03	0.950 ± 0.04
Sensitivity	0.891 ± 0.16	0.893 ± 0.15	0.914 ± 0.23	0.852 ± 0.12	0.921 ± 0.68
Specificity	0.953 ± 0.32	0.972 ± 0.04	0.933 ± 0.08	0.953 ± 0.06	0.962 ± 0.84
F Measure	0.843 ± 0.07	0.884 ± 0.06	0.724 ± 0.45	0.894 ± 0.07	0.921 ± 0.06
AUC	0.920 ± 0.27	0.930 ± 0.28	0.920 ± 0.27	0.900 ± 0.31	0.940 ± 0.33
CHG IR# (%)	↓ 51.663	↓ 30.956	↓ 25.587	↓ 29.112	↓ 18.908
CHG AUC# (%)	↑ 19.636	↑ 2.876	↑ 0.988	↑ 25.174	↑ 16.481

Note: #Changes of IR or AUC.

Table 7. Imbalanced Ratios of Subject-Specific Classifications in Test Sets.

Perform. Measures	Layer 1 $W:S$	Layer 2 REM:NREM	Layer 3 $N1$:non-$N1$	Layer 3 $N2$:non-$N2$	Layer 3 $N3$:non-$N3$
IR	7.725	6.913	5.405	1.453	2.766
Acc (%)	0.633 ± 0.24	0.793 ± 0.22	0.802 ± 0.22	0.566 ± 0.18	0.674 ± 0.22
Sensitivity	0.772 ± 0.33	0.903 ± 0.29	0.911 ± 0.28	0.722 ± 0.39	0.811 ± 0.34
Specificity	0.754 ± 0.33	0.892 ± 0.29	0.882 ± 0.28	0.702 ± 0.39	0.799 ± 0.34
F Measure	0.711 ± 0.30	0.843 ± 0.26	0.852 ± 0.26	0.604 ± 0.32	0.744 ± 0.29
AUC	0.769 ± 0.32	0.904 ± 0.31	0.920 ± 0.28	0.900 ± 0.34	0.940 ± 0.33
CHG IR# (%)	↑ 106.883	↑ 44.836	↑ 34.386	↑ 41.068	↑ 23.317
CHG AUC# (%)	↓ 16.413	↓ 2.796	↓ 0.978	↓ 20.111	↓ 14.149

Note: #Changes of IR or AUC.

The hypotheses were set as follows:

μ_1 is the mean of AUC in a sleep stage of the MLHM
μ_2 is the mean of AUC in a sleep stage of a selected model

$$\text{Null hypothesis :} \quad H_0 : \mu_1 - \mu_2 = 0,$$
$$\text{Alternative hypothesis :} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Paired-sample t-tests between the MLHM and selected models are evaluated and clearly marked in Table 5. In the case of overall SSC, there is statistically significant difference between the mean of AUC of the MLHM and a corresponding classification model at a p-value of 0.05 with the highest AUC. Statistically, MLHM is significantly different from all of the SVM in W. In R, MLHM performs better than most selected baseline ML techniques except SVM(5ep). In $N1$, MLHM has no difference with DT(10ep) and DT(30ep). SVM(15ep) solely gains the highest AUC in $N2$. In $N3$, MLHM, DT (5ep) and DT(30ep) achieve the same level of AUC.

DISCUSSIONS

This study developed the MLHM classification model, a hybrid model with multi-layer classifications. The MLHM was constructed from two baseline ML techniques, DT and SVM with specific epoch lengths. The MLHM can be used in ASSC to classifying sleep stages. A comparison of classification results between the subject-independent and the subject-specific classifications was presented. The overall classification performance of the subject-specific classification was lower than the subject-independent classification performance. In this study, we found that the ISRUC-Sleep Subgroup I is an imbalanced dataset. This could affect the classification results, therefore, a step further investigation was taken to identify the degree of imbalance. The Imbalance Ratio (IR) is one of the measures to calculate the level of imbalance. IR is a ratio between the number of instances in the majority class and the number of instances in the minority class.[33] In the case of class imbalance, AUC was suggested to be the main classification measure for our study.[32,34–36]

Table 6 and 7 show a comparison of average IRs of subject-specific and subject-independent classifications in test sets. The average IRs were calculated from 100 subjects in the dataset. In addition, the changes of IR (CHG IR) affected the changes of AUC (CHG AUC) between subject-specific and subject-independent classifications. For example, in Layer I (W:S), the IR in

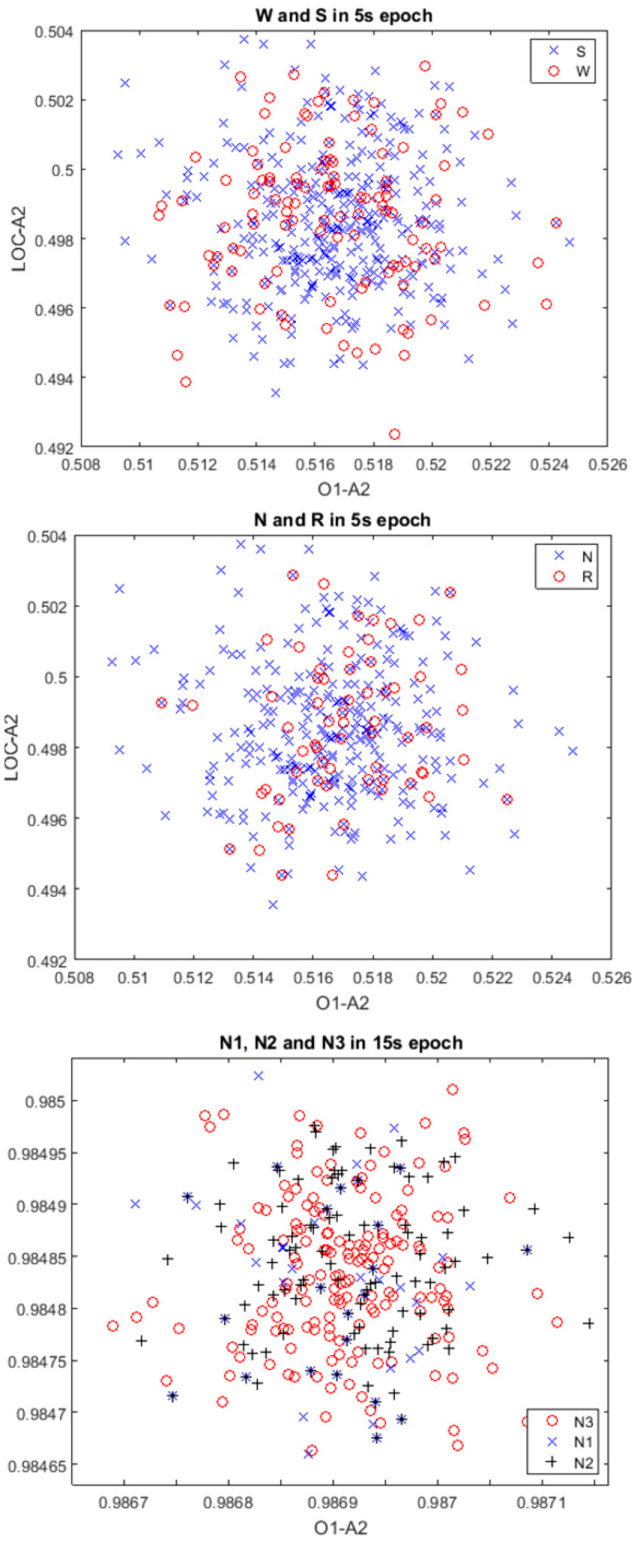


Fig. 8 Data distribution in O1–A2 LOC–A2.

subject-independent is 3.734 and the IR in subject-specific is 7.725, which is doubled. The IR is changed from subject-independent to the subject-specific by 106.883%. In terms of AUC, it drops from 0.920 to 0.769,

which is 16.413%. In Layer II (NREM:REM), the IR in subject-independent is 4.773 and the IR in subject-specific is 6.913. The IR is changed from subject-independent to the subject-specific by 44.836%. In terms of AUC, it drops from 0.930 to 0.904, which is 2.796%. Therefore, the changes of IR is inversely variated with the CHG AUC.

The MLHM was a hybrid classification model in the multi-layer architecture. The binary classification was used in the classification of W/S (Layer 1) and N/R (Layer 2). The multi-class classification was used in the $N1/N2/N3$ (Layer 3). In each of the layers, the selected classifiers acted as a filter for specific sleep stages. The classifiers were selected based on the measures especially the AUC. DT(5ep), SVM(5ep) and DT(15ep) were selected in Layers I, II and III, in the MLHM, respectively. Data visualizations of 1000 randomly selected data in O1–A2 and LOC–A2 are presented in Fig. 8. Figure 8 shows three graphical representations of data distribution in O1–A2 and LOC–A2. The top, middle and bottom graphs show the data visualization of W(red) and S(blue), N(blue) and R(red) and $N1$(blue), $N2$(black) and $N3$(red) in 5, 5, and 15 s epoch lengths, respectively. Overlapping can be visually observed. In the case of N(blue) and R(red), in the case of high degree of overlapping, the SVM was more superior in the classification than the DT. Table 3 affirmed that the best selected classifier as SVM(5ep) followed by DT(5ep). In the case of N1(blue), N2(black) and N3(red), the DT(15ep), performed the multi-class classification, was the best classifier in the list in Table 3. The 15 s epoch length shows some degree of overlapping between $N1$ and $N2$. Therefore, in Tables 6 and 7, the classification of $N3$ (AUC $= 0.940 \pm 0.33$) was better than the classification of $N1$ (AUC $= 0.920 \pm 0.28$) and $N2$ (AUC $= 0.900 \pm 0.34$).

CONCLUSIONS

SSC is one of the fundamental methods to classify sleep disorders. ASSC has been proposed over the years. ASSC usually constructed from statistical methods and ML techniques. In this study, we developed MLHM that was a hybrid classification model equipped with a newly proposed multi-layer architecture. Two baseline ML techniques included in the MLHM were DT and SVM. Specific epoch lengths were investigated and properly selected for the model. Experiments were designed to perform in both subject-specific and subject-independent classifications. The MLHM gained acceptable classification results comparing to other techniques. In-depth discussion of the classifier and epoch length selection for the MLHM were presented.

REFERENCES

1. Jafari B, Mohsenin V, Polysomnography, *Clin Chest Med* **31**:287–297, 2010.
2. Haba-Rubio J, Krieger J, *Introduction to Modern Sleep Technology*, Springer, Netherland, 2012.
3. American Academy of Sleep Medicine (AASM), Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force, *Sleep* **22**:667–689, 1999.
4. Berry RB, *Fundamentals of Sleep Medicine*, Elsevier Saunders, Philadelphia, 2012.
5. Sen B, Peker M, Novel approaches for automated epileptic diagnosis using FCBF feature selection and classification algorithms, *Turk J Elect Eng Comput Sci* **21**:2092–2109, 2013.
6. Bao FS, Lie DY, Zhang Y, A new approach to automated epileptic diagnosis using EEG and probabilistic neural network, *IEEE Int Conf Tools with Artificail Intelligence ICTAI08*, 2008, pp. 1–5.
7. Sezer E, Isik H, Saracoglu E, Employment and comparison of different Artificial Neural Networks for epilepsy diagnosis from EEG signals, *J Med Syst* **36**:347–362, 2012.
8. Kannathal N, Choo M, Acharya U, Sadasivan P, Entropies for detection of epilepsy in EEG, *Comput Methods Prog Biomed* **80**:187–194, 2005.
9. Albayrak M, Koklukaya E, The detection of an epileptiform activity on EEG signals by using data mining process, *e-J New World Sci Acad* **4**:227–250, 2009.
10. Subasi A, EEG signal classification using wavelet feature extraction and a mixture of expert model, *Expert Syst Appl* **32**:1084–1093, 2007.
11. Brignol A, Al-ani T, Drouot X, Phase space and power spectral approaches for EEG-based automatic sleep–wake classification in humans: A comparative study using short and standard epoch lengths, *Comput Methods Prog Biomed* **109**:227–238, 2013.
12. Aboalayon KAI, Faezipour M, Multi-class SVM based on sleep stage identification using EEG signal, *Health Innovations and Point-of-Care Technologies Conf*, pp. 181–184, 2016.
13. Sousa T, Oliveria D, Khalighi S, Pires G, Nunes U, Neurophysiologic and statistical analysis of failures in automatic sleep stage classification, *Int Conf Bio-Inspired Systems and Signal Processing*, pp. 1–6, 2012.
14. Fonseca P, Long X, Radha M, Haakma R, Aarts RM Rolink J, Sleep stage classification with ECG and respiratory effort, *Physiol Meas* **36**:2027–2040, 2015.
15. Srinivasan V, Eswaran C, Sriraam N, Artificial neural network based epileptic detection using time domain and frequency domain features, *J Med Syst* **29**:647–660, 2005.
16. Ozsen S, Classification of sleep stages using classdependent sequential feature selection and artificial neural network, *Neural Comput Appl* **23**:1239–1250, 2013.
17. Bharadwaj A, Minz S, Hybrid approach for classification using support vector machine and decision tree, *Int Conf*

Advances in Electronics, Electrical and Computer Science Engineering (EEC 2012), pp. 337–341, 2012.
18. Lajnefa T, Chaibia S, Rubyb P, Aguerab PE, Eichenlaubc JB, Sameta M, Kachouria A, Jerbi K, Learning machines and sleeping brains: Automatic sleep stage classification using decision tree multi-class support vector machines, *J Neurosci Methods* **250**:94–105, 2015.
19. Khalighi S, Sousa T, Pires G, Nunes U, Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels, *Expert Syst Appl* **40**:7046–7059, 2013.
20. Wongsirichot T, Hanskunatai A, A comparative investigation of PSG signal patterns to classify sleep disorders using machine learning techniques, *Lect Notes Comput Sci* **9225**:510–521, 2015.
21. Carbone T, Marrero LC, Weiss J, Hiatt M, Hegyi T, Heart rate and oxygen saturation correlates of infant apnea, *J Perinatol* **19**:44–47, 1999.
22. Sun J, Li X, Guo J, Han F, Zhang H, Identification of obstructive sleep apnea syndrome by ambulatory electrocardiography: Clinical evaluation of time-domain and frequency-domain analyses of heart rate variability in Chinese patients, *Cell Biochem Biophys* **59**:165–170, 2011.
23. Almazaydeh L, Faezipour M, Elleithy K, A neural network system for detection of obstructive sleep apnea through SpO_2 signal features, *Int J Adv Comput Sci Appl* **3**:7–11, 2012.
24. Azarbarzin A, Snoring sounds' statistical characteristics depend on anthropometric parameters, *J Biomed Sci Eng* **5**:245–254, 2012.
25. Wongsirichot T, Iad-ua N, Wibulkit J, A snoring sound analysis application using K-mean clustering method on mobile devices, *Adv Intel Syst Comput* **403**:789–796, 2016.
26. Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, Stradling J, Clifford G, SleepAp: An automated obstructive sleep apnoea screening application for smartphones, *Comput Cardiol* **40**:325–331, 2013.
27. Khalighia S, Sousaa T, Santosb JM, NunesaU, ISRUC-Sleep: A comprehensive public dataset for sleep researchers, *Comput Methods Progr Biomed* **124**:180–192, 2016.
28. Sousa T, Cruz A, Khalighi S, Pires G, Nunes U, A two-step automatic sleep stage classification methods with dubious range detection, *Comput Bio Med* **59**:42–53, 2015.
29. Khalighi S, Sousa T, Nunes U, Adaptive automatic sleep stage classification under covariate shift, *Int Conf IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2259–2262, 2012.
30. Khalighi S, Sausa T, Oliveira D, Pires G, Nunes U, Efficient feature selection for sleep staging based on maximal overlap discrete wavelet transform and SVM, *Annual Int Conf IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3306–3309, 2011.
31. Ramaswami M, Bhaskaran R, A study of feature selection techniques in educational data mining, *J Comput* **1**:7–11, 2009.
32. Han J, Kamber M, Pei J, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, Waltham, Massachusetts, 2011.
33. Garca JSV, Mollineda R, On the effectiveness of preprocessing methods when dealing with different levels of class imbalance, *Knowl Based Syst* **25**:13–21, 2012.
34. Simundic AM, Measures of diagnostic accuracy: Basic definitions, *J Int Feder Clin Chem Laboratory Med* **19**:203–211, 2009.
35. Chawla N, Japkowicz N, Kotcz A, Imbalanced data sets, *ACM SIGKDD Explorations Newsletter — Special Issue on Learning from Imbalanced Datasets*, Vol. 6, pp. 1–6, 2004.
36. Ye Y, Yang K, Jiang J, Ge B, Automatic sleep and wake classifier with heart rate and pulse oximetry: Derived dynamic time warping features and logistic model, *Systems Conf (SysCon)*, pp. 1–6, 2016.

Author Biography

Name	Mr. Thakerng Wongsirichot
Date of Birth	22 February 1981
Address	311/21 Moo 6, Thung Yai, Hat Yai, Songkhla, 90110
Education	2003 Master of Information Systems (with Distinction), University of Wollongong, Australia
	2001 Bachelor of Commerce (Electronic Commerce and Business Information System), University of Wollongong, Australia

Academic Publications

1. T. Wongsirichot and A. Hanskunatai, “A comparative investigation of PSG signal patterns to classify sleep disorders using machine learning techniques,” in Lecture Notes in Computer Sciences, vol. 9225, 2015, pp. 510-521.
2. T. Wongsirichot and A. Hanskunatai, “A Classification of Sleep Disorders with Optimal Features using Machine Learning Techniques,” *Journal of Health Research,* vol. 31, no. 3, pp. 209-217, 2017.
3. T. Wongsirichot and A. Hanskunatai, “A Multi-Layer Hybrid Machine Learning Model for Automatic Sleep Stage Classification,” *Biomedical Engineering: Applications, Basis and Communications,* vol. 31, no. 6, pp. 1-13, 2018.